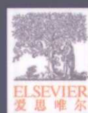


国际信息工程先进技术译丛

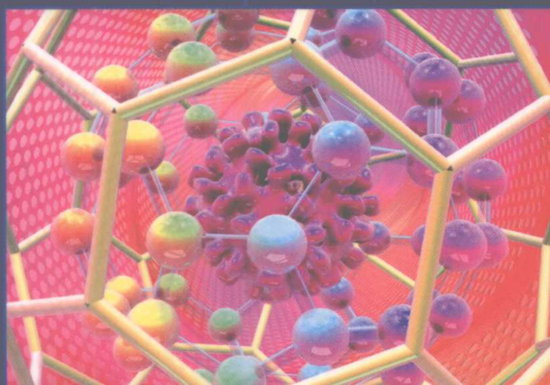


生物医学工程学概论（原书第2版）

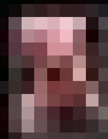
Introduction to Biomedical Engineering (Second Edition)

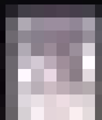
（美） John D.Enderle
Susan M.Blanchard
Joseph D.Bronzino 著
封洲燕 译





机械工业出版社
CHINA MACHINE PRESS





生物医学工程学概论（原书第2版）

Introduction to Biomedical Engineering (Second Edition)





国际信息工程先进技术译丛

生物医学工程学概论

(原书第2版)

Introduction to Biomedical Engineering

(Second Edition)

John D. Enderle
(美) Susan M. Blanchard 著
Joseph D. Bronzino

封洲燕 译

机 械 工 业 出 版 社

本书是一部全面介绍生物医学工程学的教科书，主要内容分成三大部分：生物医学工程学发展的历史背景以及职业状况，生物医学工程技术所涉及的伦理道德问题、解剖学和生理学基础知识，生物医学工程各个领域的核心知识，包括生物力学、康复工程、生物材料学、组织工程学、生物仪器、生物传感器以及生物信号处理等；各种生物医学技术应用，包括生理建模、生物信息学、计算生物学、医学成像、光学与激光在生物医学中的应用等。另外，本书的附录还介绍了计算机软件工具 MATLAB 和 SIMULINK 的使用方法。书中各章都包含了例题、习题和推荐阅读资料目录，有利于读者深入理解和掌握基本知识，并提高学生应用理论知识解决实际问题的能力。

本书不仅是生物医学工程专业难得的一本好教材，而且，对于生物学、医学和其他工程专业的学生也是一本很有价值的参考书，此外，它对于从事生物医学工程技术工作和管理工作的专业人员是一本值得常备的工具书。

Translation from the English language edition:

Introduction to Biomedical Engineering (Second Edition)/edited by John D. Enderle, Susan M. Blanchard, Joseph D. Bronzino.

ISBN: 9780122386626

本书版权登记号：图字 01-2008-4795

图书在版编目（CIP）数据

生物医学工程学概论：原书第 2 版/（美）安德勒（Enderle, J. D.），（美）布兰查（Blanchard, S. M.），（美）布及诺（Bronzino, J. D.）著；封洲燕译. —北京：机械工业出版社，2010.5

（国际信息工程先进技术译丛）

书名原文：Introduction to Biomedical Engineering (Second Edition)

ISBN 978-7-111-30125-7

Ⅰ.①生… Ⅱ.①安…②布…③布…④封… Ⅲ.①生物医学工程学概论 Ⅳ.①R318

中国版本图书馆 CIP 数据核字（2010）第 045991 号

机械工业出版社（北京市百万庄大街 22 号　邮政编码 100037）

策划编辑：牛新国　责任编辑：朱　林

版式设计：霍永明　责任校对：陈延翔

封面设计：马精明　责任印制：乔　宇

北京汇林印务有限公司印刷

2010 年 7 月第 1 版第 1 次印刷

169mm×239mm・52.75 印张・1151 千字

0001-3000 册

标准书号：ISBN 978-7-111-30125-7

定价：98.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

电话服务	网络服务
社服务中心：(010)88361066	门户网：http://www.cmpbook.com
销 售 一 部：(010)68326294	教材网：http://www.cmpedu.com
销 售 二 部：(010)88379649	
读者服务部：(010)68993821	封面无防伪标均为盗版

译 者 序

生物医学工程学是工程技术与生物学和医学高度结合所产生的一门新兴交叉学科，它旨在运用工程技术手段解决生命科学研究以及医学诊断和治疗中存在的问题，提高人类防治疾病的能力，保障人类健康。国际上，美国在第二次世界大战之后的20世纪50年代最早建立了生物医学工程学科；在我国，浙江大学于1977年最早创建了生物医学工程专业。近年来，随着我国科学技术水平的不断提高和经济建设的飞速发展，人们对于自身健康水平和生活质量的要求越来越高，生物医学工程行业也迅速发展起来。为了适应该领域的人才需求，国内已有百所以上高校相继开设了生物医学工程专业。但是，生物医学工程学包罗万象，几乎涉及所有工程学科以及生物学和医学，目前能够全面而系统地讲述该专业的理论知识和技术应用的导论性教材比较缺乏。

本书英文版由31位学者撰写，可以说是目前国际上覆盖生物医学工程各个领域最全面的一部教材，信息量相当大，理论与应用紧密结合。从医疗技术自古至今的发展历史，到先进科学技术应用所引发的伦理道德问题；从各种先进技术的基本原理，到美国食品与药物管理局对于医疗器械产品的管理和论证制度；从众多领域的严谨的理论知识，到适合于学生思考和练习的深入浅出的例题和习题；本书向读者展现了生物医学工程丰富多彩、引人入胜的专业景象。除了有关本学科的发展背景、伦理道德问题以及基础生物医学知识以外，为了使脉络清晰，本书将其他内容分成核心理论和应用技术两大部分，这体现了作者的一种尝试。正如作者在前言中所述，即便如此，本书仍然没能囊括生物医学工程现有的所有领域，并且，科学技术的进步还将不断给该学科增添新的内容，健康和幸福永远是人类的奋斗目标。希望本书的翻译能够给生物医学工程专业的学生和从业人员提供一本有价值的专业著作。

本书英文版中有涉及一些宗教和道德伦理观点的内容，并不代表机械工业出版社及译者的观点；此外，第2章论述的一些医学道德与伦理问题，目前越来越受到社会的广泛关注，是颇有争议和值得深思的，书中所表述的各种观点仅供读者参考。

翻译过程中，对原书某些明显的笔误或印刷错误已作了更正，为保持译著的简洁流畅，有些没有加以标注；同时为尽量保持原书特色，书中部分图形和文字符号并未按国家标准修改，请读者注意。

本书的翻译得到了国家自然科学基金项目的资助（项目编号为30770548和30970753）。郑晓静、吴丹、汪洋、杨彭举、田聪、毛盾、徐白露、李林森、蓝义昀、李迎等同学参加了本书的大量初译工作，在此深表谢意。

本书内容所涉及的研究领域极其广泛，限于译者的水平，翻译过程中难免存在错误和不妥之处，恳请广大读者批评指正。译者电子信箱：hnfzy@yahoo. com. cn。

封洲燕

浙江大学　生物医学工程与科学仪器学院

2010年2月

前　言

本书第 2 版的目的与第 1 版相同，也就是全面地概述生物医学工程学科。新版许多章末尾添加了习题，其中有几章被合并在一起，有几章则被全部删除，另外增添了几章反映该学科最新发展的内容。

在过去 50 年的发展历程中，生物医学工程学已经清楚地展现出其多种多样、包罗万象的特性。本教材包括了生物电、生物信息学、生物材料学、生物力学、生物仪器、生物传感器、生物信号处理、生物技术、计算生物学和复杂性、基因组学、医学成像、光学与激光、放射成像、康复工程、组织工程、道德与伦理问题等内容。显然，本书不可能囊括生物医学工程学科的所有研究领域，但是，我们尽可能包含了其大部分主要领域。

本教材主要面向已经学习过微积分和静力学基础的工程专业学生，大学二年级和三年级的学生都应该已经具备学习本书内容的基础知识，生物学、医学和护士专业的学生如果掌握了所需的数学基础知识，也可以读懂本书。

虽然在本书的论述和证明过程中我们都力求严谨，但我们的最终目标是帮助学生掌握生物医学工程学的本质内容。因此，为了便于理解，必要时我们偶尔会在数学的严谨上做出让步。为了解释基本概念、培养学生解决问题的能力，书中列举了大量例题，并使用计算机软件 MATLAB 和 SIMULINK 作为工具帮助解题。SIMULINK 是 MATLAB 的一个工具箱，可用于仿真动态系统。本书的附录介绍了 MATLAB 和 SIMULINK 的基本使用方法，该软件工具是 Mathworks 公司的产品，公司联系方式为

The Mathworks，Inc.
24 Prime Park Way
Natick，Massachusetts 01760
Phone：(508) 647-7000
Email：info@mathworks. com
WWW：http：//www. mathworks. com

本书各章的内容分别包含了生物医学工程某个特定领域的主要发展历史，以及该领域生物医学工程设计、分析和建模过程的基本原理，并且，举例介绍了该领域所遇到的某些问题及其解决方法。另外，按照正文内容的顺序在各章末尾选列了一些习题，难易皆有。

根据教师的需要和兴趣，本教材可以安排为半学年、一学年或者四分之三学年的课程设置。书中第 1 章介绍生物医学工程学发展的历史背景和职业状况；第 2 章介绍有关生物医学伦理道德的重要问题；第 3 章讲述解剖学和生理学基础知识；第 4～

10 章讲述生物医学工程各个领域的核心知识，包括生物力学、康复工程、生物材料学、组织工程学、生物仪器、生物传感器以及生物信号处理。其中生物仪器、生物传感器和生物信号处理这几章内容之间的相互联系较紧密，建议教师在安排课程计划时将它们组织在一起讲解。本书后面的第 12～17 章讲述各种生物医学工程技术。

本书的勘误表等材料请参见网址 http：//intro-bme-book. bme. uconn. edu/。

致谢

本书的撰写工作得到了许多人的帮助。首先要感谢各章作者在非常紧张的时间里完成了书稿；其次，要特别感谢 Elsevier 出版社，尤其要感谢 Christine Minihane 和 Shoshanna Grossman 两位编辑的辛勤工作；另外，还要感谢项目管理人 Karen Forster，以及指导出版过程的 Kristin Macek。

衷心感谢本书第 1 版编辑 Joel Claypool 以及 Elsevier 科学出版社的 Diane Grossman，从最初 1996 年在 Amsterdam 喝咖啡时的一次谈话，到 2000 年本书第 1 版的出版，他们为此付出了巨大的努力。

原书第 1 版作者名单

Susan M. Blanchard
Florida Gulf Coast University
Fort Myers, Florida

Joseph D. Bronzino
Trinity College
Hartford, Connecticut

Stanley A. Brown
Food and Drug Administration
Gaithersburg, Maryland

Gerard Coté
Texas A&M University
College Station, Texas

Roy B. Davis III
Shriners Hospital for Children
Greenville, South Carolina

John D. Enderle
University of Connecticut
Storrs Connecticut

Robert J. Fisher
University of Massachusetts
Amherst, Massachusetts

Carol Lucas
University of North Carolina
Chapel Hill, North Carolina

Amanda Marley
North Carolina State University
Raleigh, North Carolina

Yitzhak Mendelson
Worcester Polytechnic Institute
Worcester, Massachusetts

Katharine Merritt
Food and Drug Administration
Gaithersburg, Maryland

H. Troy Nagle
North Carolina State University
Raleigh, North Carolina

Joseph Palladino
Trinity College
Hartford, Connecticut

Bernhard Palsson
University of California at San Diego
San Diego, California

Sohi Rastegar
National Science Foundation
Arlington, Virginia

Daniel Schneck
Virginia Polytechnic Institute & State University
Blacksburg, Virginia

Kirk K Shung
Pennsylvania State University
University Park, Pennsylvania

Anne-Marie Stomp
North Carolina State University
Raleigh, North Carolina

Andrew Szeto
San Diego State University
San Diego, California

LiHong Wang
Texas A&M University
College Station, Texas

Steven Wright
Texas A&M University
College Station, Texas

Melanie T. Young
North Carolina State University
Raleigh, North Carolina

原书第 2 版作者名单

Susan M. Blanchard
Florida Gulf Coast University
Fort Myers, Florida

Joseph D. Bronzino
Trinity College
Hartford, Connecticut

Stanley A. Brown
Food and Drug Administration
Gaithersburg, Maryland

Gerard Coté
Texas A&M University
College Station, Texas

Charles Coward
Drexel University
Philadelphia, Pennsylvania

Roy B. Davis
Shriners Hospital for Children
Greenville, South Carolina

Robert Dennis
University of North Carolina
Chapel Hill, North Carolina

John D. Enderle
University of Connecticut
Storrs, Connecticut

Monty Escabí
University of Connecticut
Storrs, Connecticut

R.J. Fisher
University of Massachusetts
Amherst, Massachusetts

Liisa Kuhn
University of Connecticut Health Center
Farmington, Connecticut

Carol Lucas
University of North Carolina
Chapel Hill, North Carolina

Jeffrey Mac Donald
University of North Carolina
Chapel Hill, North Carolina

Amanda Marley
North Carolina State University
Raleigh, North Carolina

Randall McClelland
University of North Carolina
Chapel Hill, North Carolina

Yitzhak Mendelson
Worcester Polytechnic Institute
Worcester, Massachusetts

Katharine Merritt
Food and Drug Administration
Gaithersburg, Maryland

Spencer Muse
North Carolina State University
Raleigh, North Carolina

H. Troy Nagle
North Carolina State University
Raleigh, North Carolina

Banu Onaral
Drexel University
Philadelphia, Pennsylvania

Joseph Palladino
Trinity College
Hartford, Connecticut

Bernard Palsson
University of California at San Diego
San Diego, California

Sohi Rastegar
National Science Foundation
Arlington, Virginia

Lola Reid
University of North Carolina
Chapel Hill, North Carolina

Kirk K. Shung
Pennsylvania State University
University Park, Pennsylvania

Anne-Marie Stomp
North Carolina State University
Raleigh, North Carolina

Tom Szabo
Boston University
Boston, Massachusetts

Andrew Szeto
San Diego State University
San Diego, California

LiHong Wang
Texas A&M University
College Station, Texas

Stephen Wright
Texas A&M University
College Station, Texas

Melanie T. Young
North Carolina State University
Raleigh, North Carolina

目　　录

第 1 章　生物医学工程学的发展历史

Joseph Bronzino 博士，专业工程师

本章目录

本章主要内容：

- 先进医疗技术在现代化医疗保健体系创建过程中所发挥的作用；
- 生物医学工程的定义和生物医学工程师在医疗保健体系中的作用；
- 为什么生物医学工程师属于专业人员。

工业化国家的技术发展日新月异，几乎已经渗透到我们生活的各个方面，尤其是医学和医疗卫生领域。虽然医学的发展历史悠久，但是，相对而言，真正能够广泛有效地提供诊断和治疗服务的高科技医疗保健体系却还属于新兴事物，其建立比较晚。并且，在其发展过程中，拥有先进技术的医疗保健中心——现代化医院的出现具有特别重要的意义。

工程技术对于先进医疗保健体系的发展具有巨大的推动作用，大批工程人员投入到各种医疗创新的研究工作中，使得医学和工程学这两个充满活力的学科结合在一起，从而形成了生物医学工程学。该学科的从业人员开发各种新技术，比如生物传感器、生物材料、图像处理和人工智能等，用于医疗保健领域的研究、诊断和治疗，帮助人类攻克各种疾病。

生物医学工程专业人员属于医疗保健队伍中的新成员，他们的职责是研发新技术，解决现代医疗保健领域所面临的各种疑难问题。本章将全面概述工程技术在现

代医疗保健体系形成过程中所起的作用，介绍生物医学工程人员的基本职能，并且展示这个生机勃勃的新兴领域的职业状况。

1.1 现代医疗保健体系的形成过程

远古时期，人们凭借本能和经验，创建了最早的医药技术。例如，通过收集和编撰一些可靠的医疗案例，人们开始建立草药治病法、正骨疗法、外科手术和助产术等。就像原始人类经过长期观察总结出某些蔬菜和谷物可以食用并种植一样，人们经过长期的考察，也掌握了某些疾病的特性，并把这些经验传给后代。

有证据表明，人类很早之前就积极地利用各种工具施行外科手术、治疗疾病。例如，在欧洲、亚洲和南美洲的许多地方，人们发现了被类似环钻的工具开过孔的颅骨。显然这些骨头上的孔都是用硬器开出来的，构成了进出大脑的通道。那么，当时为什么要做这种手术呢？现在人们只能进行猜测，最有可能的还是，如果有人偏头痛剧烈发作，或者癫痫病人病情突发并瘫倒在地，那么，当时的人们也许认为这种开颅手术可以解救病人。而且，这些头骨上的孔的边缘是圆滑的，说明手术后骨头又生长过。这点证明了手术是在活人身上进行的，这些人术后确实活了下来。从这些早期医术开始，医学职业就成了整个人类社会文化不可分割的一部分。

下面介绍几位最成功的早期行医者，他们的故事都很有趣。例如，长期以来，埃及人一直极为敬重的 Imhotep，其实就是公元前 3000 年设计第一座金字塔的建筑师。不过，他并不是因金字塔闻名，而是以医生的身份闻名遐迩。在给病人看病时，他可以使病人安详地入睡，Imhotep 这个名字就是由此而来，它象征“带来安宁的人”。Imhotep 的医术非常精湛，以至于他在埃及文化中被神化为“康复之神”。

又例如，至今仍在广泛使用的医药处方的标志符“Rx”，就来源于古埃及的传说“太阳神 Horus 的眼睛”（Eye of Horus）。Horus 小时候曾遭受恶魔 Seth 的袭击而失明，后来，他母亲 Isis 向最权威的健康之神 Thoth 求助，Thoth 随即恢复了 Horus 的视力。于是，“Horus 的眼睛”在埃及成为神灵保佑和康复的象征，由此传说产生的“Rx”标志就成了古代医学和现代医学之间的最明显的纽带。

Imhotep 的思想和实践经验，被许多人记录在莎草纸上，并埋藏在古墓中。1873 年，George Elbers 得到了一卷大约写于公元前 1500 年的书稿，其中记录了从治疗鳄鱼咬伤到便秘等各种疾病的几百种方法。在此之前，1862 年，Edwin Smith 曾经找到另一卷大约写于公元前 1700 年的莎草纸书稿，它是最重要最完整的一本古代外科手术著作。这两本莎草纸著作描述了一系列手术病例的诊断、预后和治疗方法。毫无疑问，它们是医学史上的杰出著作。

受到古埃及 Imhotep 的影响，希腊人树立了自己的康复之神 Aesculapius。据说，太阳神 Apollo 在一次下访人间时留下了一个取名为 Aesculapius 的儿子。Apollo 是一位非常关心子女的父亲，正如现今很多父母一样，他想使自己的儿子成为一名医生，于是就请人首马身的 Chiron 当老师，教 Aesculapius 学习治病的方法。不久，学生不

仅精通了老师的医术，还迅速超过了老师，使人们的寿命大大延长，以至于阴间的人口数量都开始下降了。于是，阴间之神 Pluto 提出了强烈的抗议，迫使天神 Zeus 不得不用雷电劈死了 Aesculapius，并将他升至天堂成为神。

神话故事一般都与史实具有一定的联系，Aesculapius 究竟是不是像埃及的 Imhotep 一样，是一位人间行医者呢？这无从考证。不过，有一件事是可以确定的，那就是到了公元前 1000 年，医生已经成为十分受人尊敬的职业。用于康复治疗的古希腊 Aesculapius 教堂可以说是最早的医院（见图 1.1）。

古希腊医学的代表人物希波克拉底（Hippocrates）出生于 Cos 岛，岛上就有一座“康复”教堂，并且是这类教堂中最有名的一个。希波克拉底从小随父亲学医，他并不是一位很善于创新的医师，但是他很善于收集和学习前人的各种治疗方法和技术。他把医师看成科学家，而不是神父，因此，希波克拉底最早给医学注入了一种新的基本要素，这就是科学精神。从他开始，行医时的迷信仪式逐渐变成了检查、诊断和临床治疗。他认为疾病是一种自然发生的过程，而不是神赐予的。疾病的发展有一定的规律，病症是身体对抗疾病的反应。他还强调身体本身拥有自我康复的能力，医师的任务只是帮助增强这种自然抗病的力量。希波克拉底把每个病人都当做新的案例来研究，并将它们记录下来。他对各种疾病的精确描述至今仍是医生们学习的楷模。希波克拉底曾任 Cos 学院的院长，该学院培养了很多学生，这些人后来移民到了地中海各个地区，在那里行医并广泛传播他们院长的哲学思想。希波克拉底及其学院的工作，以及他所创立的传统，造成了医学与早期巫术之间的第一次真正决裂，并为科学医学打下了基础。不过，作为一名医师，希波克拉底代表的是医学精神，并不是医学科学，他体现了一名好医师的形象，即病人的朋友、仁慈的专家。



图 1.1　被送进 Aesculapius 教堂的患病儿童（摘自 http：//www.nouveaunet.com/images/art/84.jpg）

在古罗马帝国的鼎盛时期，其威力遍及大半个地球。它吸收并继承了各种伟大文化，其中包括医学上的进步。虽然罗马人自己对临床医学的发展并没有什么贡献，但他们在公共卫生方面确实取得了杰出的成就。例如，他们建立了组织严密的军队医疗服务队。这些服务队不仅在各个战场上给古罗马军团提供急救服务，而且还在遍布罗马帝国的战略要塞为病人的康复治疗建立了基地医院。城市下水道系统和输水管道的建设是罗马人又一卓越的成就，为罗马帝国创造了卫生的生活环境，体现了其在医学和社会环境上的进步。饮用纯净水和干净食物可以控制流行病的传播，

使城市得以生存下去。不幸的是，由于对疾病缺乏足够的科学认识，使用公共卫生设施的罗马人即使想尽办法，也不能避免周期性发生的疾病灾难，特别是瘟疫，它们随时会无情地降临到市民头上。

起初，罗马统治者并不喜欢希腊的医师和医术。然而，几年之后，希波克拉底的弟子们给人们留下了良好的行医印象，并广泛流传开来。在公元前46年，古罗马凯撒大帝为了奖赏这些行医者为罗马帝国人民作出的贡献，给所有在他领地上的希腊医师们授予了罗马公民的身份。这个新身份使这些人的生活有了保障，在当年罗马遭遇饥荒时，这些希腊医师是惟一没有被驱逐出境的外国人。相反，罗马人甚至还给他们奖金让他们留下来！

颇具讽刺意味的是，被认为是罗马历史上最伟大的医师盖仑（Galen）竟然是希腊人。因为治愈了罗马帝王的热病，盖仑成为罗马医学界的名人。他傲慢，喜欢吹嘘，与希波克拉底不一样，他只报道成功的病例。不过，他的确是一位杰出的医师。盖仑的医疗诊断是一门艺术。除了负责自己的病人以外，他还答复来自远方其他地方的医疗咨询。他非常勤奋，撰写了300多部有关解剖体检的书籍，其中包括精选的历史案例、他开过的药方、当然也少不了一些自我炫耀的内容。他当时写的人体解剖学其实是蒙人的。由于他反对人体解剖，因此，只是根据动物的研究结果来画人体的解剖器官。但是，由于盖仑当时主宰了医学界，后来又得到罗马天主教的支持，因此，他实际上也阻碍了医学的求真探索。他的医学观点和著作在随后的宗教统治的“黑暗时期”被罗马教皇和权威人士看作“圣经”和“定律”。

随着罗马帝国的灭亡，教堂成了贮藏知识的地方，尤其是聚集了那些经过几个世纪的漂泊，流传到地中海地区的知识宝藏。这些知识宝藏包括医学知识，它们散布于各个不同级别的教堂，在教徒中流传。

早期天主教的教义和人们对神的恩慈的信仰使得医学研究在上千年时间里一直停滞不前。在16世纪文艺复兴时期之前，医学一直没有出现显著的进步。希波克拉底曾经指出，疾病不是上帝给予人类的惩罚，而是一种自然现象。但是，在教会和新上帝的指引下，疾病超自然起源的陈旧观点又复苏并传播开来。

在这个时期，虽然医学知识缺乏，不过，穷人却能获得慈善救助。基督徒医师们对穷人和富人一视同仁，教会承担着救死扶伤的责任。而且，现在的医院实际上就是从基督教当时创建的医院演变而来的，因此，创建医院被公认为教徒医学的主要贡献之一。公元335年，在第一个信奉基督教的罗马帝王君士坦丁一世（Constantine I）的统治下，所有异教徒的康复教堂都被关闭，并且每个主教堂城市都建立了医院。（注：英文的医院“hospital”一词源于拉丁文hospes，意思是“主人”或“招待者”。英文的“hotel”和“hostel”两词也源于该词根。）这些早期的医院都是一些简朴的房子，可以给疲惫的旅客和病人提供食宿和护理。医院由教会运作，修道士和修女们负责病人的康复治疗。

在十字军时期，基督教信仰、人道主义和慈善思想在欧洲和中东地区广泛传播，医院系统也同时传播开来。但是，那些受过培训的医生主要还是到病人家中给人看

病。医院接纳的只是那些过于劳累的游客、穷人和没有治愈希望的病人。并且，早期的医院条件差别很大，有些医院资金充足，管理妥善，对待病人友好。但是，大多数医院其实只是看守所，其作用主要是把棘手的或者患传染病的病人与公众隔离开来。这些地方往往拥挤、肮脏，病人和医护人员的死亡率都很高。因此，人们通常害怕并躲避医院。

15～16 世纪的文艺复兴和宗教改革解除了教会对医院和医疗活动的控制。文艺复兴时期，人们追求医学知识等自然界真正奥秘的渴望再次被激发。人体解剖学得到了进一步发展，并且，米开朗基罗（Michelangelo)、拉斐尔（Raphael)、杜勒(Durer）等著名画家，当然还有天才画家达·芬奇（Leonardo da Vinci)，都为深入研究人体解剖学作出了贡献。他们把盖仑用文字描述的人体画成实物图像，非常详细地描绘出病人以及他们的痛苦，并且惊人地深入勾画出心脏、肺、大脑和肌肉的结构。他们还试图在人物画像上展现出情绪和身体素质。在这个激昂奋发的时代，医师们也开始用同样的方式对待病人，并深入探索医学知识。新的医学院不断涌现，最著名的有萨勒诺（Salerno)、博洛尼亚（Bologna)、蒙彼利埃（Montpelier)、帕多瓦（Padua）和牛津（Oxford）等地的医学院。这些医学院重新采纳了希波克拉底的学说，认为病人也是人，疾病是一种自然过程，基于知识的治疗可以帮助人体战胜疾病。

文艺复兴时期，检测技术也开始发展。1592 年伽利略访问意大利城市帕多瓦(Padua）时，给一大群医学院的学生作了数学讲座，讲解他的著名理论，并展示了他的一些发明，有测温器、钟摆和望远镜镜片等。他的学生 Sanctorius 就利用这些装置研究了人体体温和脉搏。后来，帕多瓦的毕业生威廉·哈维（William Harvey）还应用伽利略的运动定律和力学定律求解了血液循环问题，测出了流经动脉的血流量，用于诊断心脏的功能。

伽利略提倡运用正确的实验测量方法帮助医生进行有效的诊断，而不是不计后果的猜测。有些医学研究人员采用这些新方法，将体温和脉搏速率的测量数据与其他相关症状结合起来，帮助医生诊断某些特殊的疾病。与此同时，显微镜的发明扩大了人类的视野，将一个惊奇的未知世界呈现在人们眼前，成为备受关注的焦点。不过，这些科学新装置当时对于普通医师并没有产生什么影响，他们仍然在使用放血法和有害的药膏治疗疾病。只有在大学里，多个科研小组联合起来，才能发挥各种智慧，汇集各种装置，发明和使用科学的新仪器。

在英格兰，亨利八世（Henry Ⅷ）曾经热情有力地支持医学，他支持医生反对渎职行为，并支持建立皇家医学院（Royal College of Physicians)。该医学院是欧洲最早的纯医学机构。16 世纪初期，他镇压了修道士体制，教会医院都被所在城市接管。随后，出现了私立的、非营利性的志愿者医院，工作人员是医生和医学院学生，而不是修女护士和修道士医师。但是，这些医院机构中几乎没有专业护理人员。只有教会医院还保留着护理，使得可怜的病人们只能留在这些公共教会医院。后来佛罗伦萨·南丁格尔（Florence Nightingale）（见图 1.2）改变了这种局面。

这里，还有一件值得一提的事件。随着城市人口的不断膨胀，英格兰对医院的需求量空前增长，但是，当时的设施不可能满足那么大的需求。因此，在17世纪，伦敦的两家主要城市医院——圣巴塞洛缪(St. Bartholomew) 和圣托马斯 (St. Thomas) 率先采取了一项措施，医院只接纳和治疗那些可能治愈的病人，不能治愈的患者只能把自己的命运交给精神病收容所、监狱或者救济院等其他机构。



图 1.2 佛罗伦萨·南丁格尔的肖像
(摘自 http: //ginnger. topcities. com/cards/computer/nurses/765x525nightengale. gif)

18世纪，在法国和美国殖民地，人道主义和民主运动成为备受人们关注的问题，权利平等的思想终于形成。并且，随着城市化的发展，美国社会开始关注其成员的福利。医护人员把服务范围扩大到那些社会的“不幸”群体上，开展监狱改革、儿童保育和医院运动，来帮助缓解这些“不幸者”的痛苦。然而，尽管医院开始积极主动地救治病人，但是，病人的死亡率并没有下降，反而继续增长。例如，巴黎的Hotel Dru 医院大约建于17世纪，是现存的最古老的医院，在1788年，该医院病人的死亡率接近25%。不仅病人的死亡率高，而且在这些医院工作的医护人员的死亡率也相当高，每年有6%～12%的医护人员死亡。

实际上，当时医院成了人们尽量回避的地方，因此早期美国殖民者迟迟不愿意建立医院，美国第一家医院——费城的宾州医院 (Pennsylvania Hospital) 直到1751年才建成。波士顿则用了200年的时间才建起了第一家麻省总医院 (the Massachusetts General Hospital)，并于1821年向公众开放。

直到19世纪，医院才有能力使相当数量的病人受益。这个时期的进步主要归功于护理水平的提高。从克里米亚战场 (Crimean War) 回到英格兰的佛罗伦萨·南丁格尔创建了护理职业，她认为当时医院的状况比疾病本身更容易引起病人的死亡。19世纪前半叶，南丁格尔迫使医院将医疗重点放在病人的护理上。满怀着激情和智慧，她阐述了护理思想：“护理可以使我们拥有可能的最佳条件，以便于病人自然恢复健康，并保持健康……请注意，我们是护理病人，而不是护理疾病。”到了19世纪后半叶，她的影响力达到了顶峰，世界各地新建的医院几乎无不采纳她的忠告。

虽然这些工作很出色，但是19世纪的医院仍然只是接纳穷困病人的机构。例如，19世纪70年代，美国人在审核约翰·霍普金斯医院 (Johns Hopkins Hospital) 的建设规划时认为，配置324张免费病床和24张收费病床是很合理的。那时，不仅医院接纳的病人只代表广大社会阶层中的很小一部分，而且医院也只能治疗各种疾

病中很有限的几种疾病。例如，1873 年，大约半数美国医院不接纳传染病病人，很多医院也不接纳不可治愈的病人。此外，在那个时期，普通医院手术率只有 5%，其中多数是外伤病例。

一个世纪以前的美国医院相当简陋，其管理机构对于医院的装备没有专门的规定，只要有炊具和洗衣设备就可以了。此外，由于医生通常没有薪水，并且护理费用很低，因此，医院正常运作的大部分费用开销在食物、药品和公用设施上。直到 20 世纪，美国才形成了现代化医院，下面我们就会看到，工程技术在医院的发展过程中起到了非常重要的作用。

1.2 现代医疗保健体系

现代医学实际上始于 19 世纪末 20 世纪初。在 20 世纪之前，由于主要医学资源只有为数不多的医生，且医生所受的教育很有限，因此，医学对于普通市民没什么贡献。当时，造成医生短缺的原因与现在不同。那时，医生收取的费用很低，需求也很少。医生所能提供的大部分服务，居住在社区里的有经验的业余人士也都能提供。治疗和康复主要是上门服务，亲戚和邻居有能力也愿意帮助护理病人，接生婴儿还有专门的助产师。而那些家庭治疗不能治愈的病人也只能听天由命。到了 20 世纪，科学技术的重大发展才促进了美国医疗保健体系的迅速发展。于是，医院成为该体系的中心，专业医生和护士成为主角。

20 世纪，化学、生理学、药理学等基础科学开始加速发展，形成了一个各学科之间紧密结合、交叉发展的氛围。物理学的各种新发现促使医学研究产生了飞跃。例如，1903 年，威廉·爱因托芬（William Einthoven）设计出了第一台心电图机，测量了心跳过程中所产生的电流变化。这一发明不仅为心血管医学，也为电测量技术开创了一个新时代。

那个时期的新发现真是层出不穷，就像联锁反应一样，其中对临床医学发展意义最大的是 X 射线的发现。从伦琴（W. K. Roentgen）阐述他的“新射线”那一天起，人体的秘密就开始向医学检查敞开了。开始时，X 射线主要用于骨折和脱臼的诊断。美国大多数城市医院都用上了 X 射线机这个新设备。于是，独立的放射科在医院里应运而生，并且其影响迅速扩大。几乎每个医疗科室，如外科、妇产科等，都借助这项新技术迅速发展。到了 20 世纪 30 年代，钡盐等各种射线不能穿透的物质的应用，使得 X 射线机几乎可以将人体所有器官可视化。

X 射线技术的发明极大地提高了医生的能力，使他们能够正确地诊断各种疾病和损伤。并且，借助这项技术，医院开始改变被动接纳穷困病人的处境，转变为能够主动积极地救治美国社会所有公民的医疗机构。

20 世纪 30 年代中期磺胺类药物的开发，以及 40 年代青霉素的开发，极大地减少了住院病人之间交叉感染的危险性。有了这些新药品，外科医生手术时也可以避免感染，不会得病和死亡。还有，虽然在 1900 年人们就已经发现了人具有不相容的

不同血型，并且在1913年就开始用柠檬酸钠来防止血液的凝固，但是，直到20世纪30年代冷藏技术成熟之后，血库才开始全面推广使用。所以，在此之前，人们献血时，都是新鲜使用，甚至还保留着献血者的体温。

随着科学技术的蓬勃发展，美国医学的威望也日益上升。1900～1929年期间，生理学和医学诺贝尔奖获得者主要来自欧洲，根本没有美国人；但是，从1930～1944年，就是第二次世界大战结束之前，有7位美国人荣获此奖项；在第二次世界大战之后的1945～1975年，有37位美国生命科学家获此殊荣；而从1975～2003年，美国人获奖人数达到了40名。自1930以来，共有79位美国科学家凭借他们的杰出研究成就荣获了诺贝尔奖。这些研究成就的绝大多数都得益于科学技术的临床应用（见图1.3）。

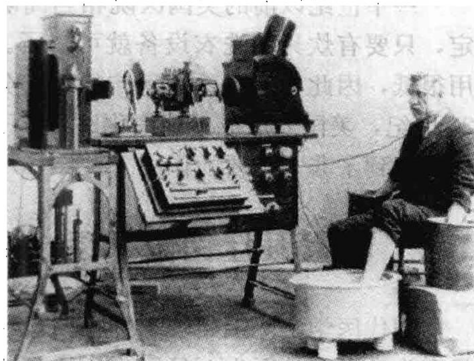
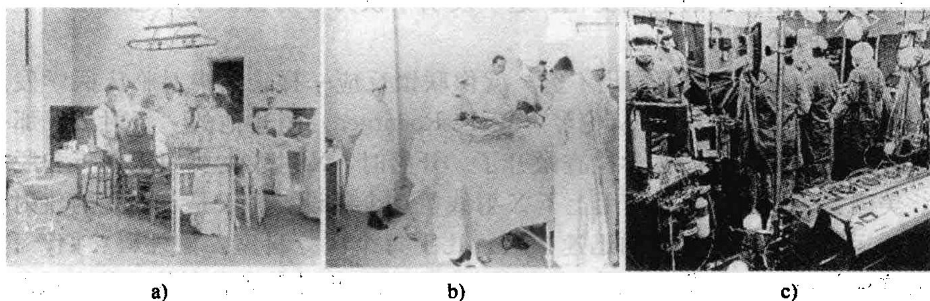


图1.3 早期心电图机的照片

科学技术的应用推动了复杂外科手术技术的快速发展（见图1.4）。1927年Drinker发明了呼吸机，1939年有人创建了心肺分流术（heart - lung bypass）。20世纪40年代，人们开发了心导管及心血管造影术（cardiac catheterization and angiography），也就是将细导管通过手臂静脉插入到心脏，然后通过导管注入不透射线的染料，从而可以在X射线照片上显示出心脏和肺部的血管和瓣膜，用于正确诊断先天性和获得性心脏病（主要是风湿热引起的心脏瓣膜损坏等）。这些技术开辟了心血管外科手术的新纪元。



a) b) c)

图1.4 手术室的变化

a）19世纪末20世纪初时手术室的场景 b）20世纪20年代末和30年代初时手术室的场景 c）如今的手术室场景（摘自JD Bronzino的"Technology for Patient Care"，Mosby：St. Louis，1977；"The Biomedical Engineering Handbook"，CRC Press：Boca Raton，FL，1995；2000；2005）

现代科学技术发展的另一个产物就是电子显微镜技术，它在20世纪50年代开始用于医学，极大地推进了微小细胞的可视化。检测肿瘤的人体扫描技术就是基于这

种原理，这类技术的应用使社会不得不进入原子时代。这种“肿瘤检测”使用放射性药物，在当时各个医院新成立的核医学科室成为常规技术。

上述发明和其他许多发明一起对医学产生了巨大的影响。随着医学不断吸收新科学技术，早先由“江湖郎中”似的医师组成的医疗保健体系已经一去不复返，取而代之的是以现代化医院为中心的体系。

第二次世界大战之后，综合性医疗保健技术加速发展，许多军用先进技术转化为民用，在这种强大的技术力量冲击下，医学界受益匪浅。这里以电子技术领域的发展为例。曾经用于跟踪敌人军舰和飞机的技术，以及给飞行员提供海拔、气流速度等信息的技术，都被用于医学领域，用来记录中枢神经系统神经元的微弱电活动，或者用于监测病人的心跳。

历史上，科学和技术这两者的发展总是交替进行的。如果要寻找两者之间的因果关系，你会发现有时技术是起因，科学是结果；而有时却是科学的进步导致了技术的形成。例如，射击产生了弹道学，蒸汽机产生了热力学，动力飞行产生了空气动力学。这些都是技术先于科学的例子。但是，随着电子学的发展，这种技术与科学之间的因果关系产生了变化，变成了系统地利用科学研究成果和科学知识来实现技术应用目标。

例如，应用人类登上月球的科学知识开发的各种仪器设备已经不计其数。在医院，监测病人心脏活动的无线遥测装置就是一例。有了它，病人和医生的活动范围都不再受到病房四周墙上拉出的电线的限制。那些用于控制阿波罗号太空舱飞行的计算机，如今在社会上已是随处可见，20 世纪 70 年代以来，这种“电脑”也已广泛用于医学领域的各种复杂计算、数据记录，甚至用于控制维持生命的仪器设备。计算机 X 射线断层扫描（Computerized Tomography，CT）和磁共振成像（Magnetic Resonance Imaging，MRI）等新兴医学成像技术的发展（见图 1.5），就完全依赖于不断进步的计算机技术。20 世纪三四十年代的许多科学幻想，现在都已经成为现实。仪器设备不断接受最新的技术革新，而许多技术又会很快过时，科学技术的发明和应用不胜枚举。

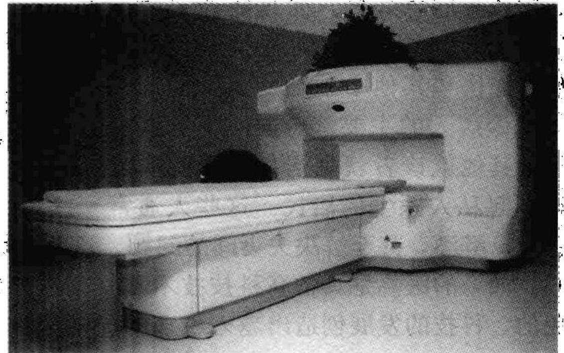


图 1.5　现代医学成像设备

（摘自 http：//137.229.52.100/～physics/p113/hasan/）

人工脏器手术现在已经属于常规手术。随着 1954 年第一例肾移植手术的成功，人造器官的设想获得了公认，并正式在医学界普及开来（见图 1.6）。人造心脏瓣膜和人造血管等假体的制造技术迅速发展，甚至已经开始开发整个人工心脏，用于取

代损坏的人体心脏。虽然至今人工心脏还不能完全令人满意，但它确实已经给一些急需的病人提供了辅助心脏的功能。这些技术的发明从根本上改变了外科手术的性质和运用。一个世纪以前，医院的外科手术还只是处于次要的地位，而如今医院的外科手术具有极其重要的作用，这种变化体现了科技发展给医疗卫生行业和医院机构带来的变革。

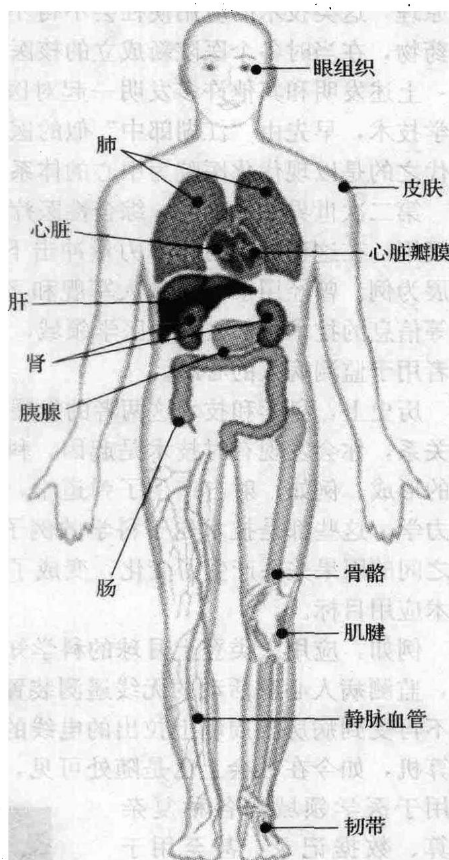


图 1.6 人体可能移植部位的示意图（摘自 http://www.transplant.bc.ca/images/what_organs.GIF）

这些变革也使现代化医院成为医疗保健体系的核心机构。由于复杂昂贵的高科技只有医院才能拥有，并且，面向培养临床医生和科研人员的医学博士生教育越来越与高科技密切相关，因此，医生和病人都逐渐被吸引到医院。另外，20 世纪五六十年代，日益加剧的分布不均和医生的明显短缺也迫使越来越多的病人和医生随时转向城市医院的救护诊所和急救监护病房。

救护诊所不仅要处理日益增加的突发事件（多数与酗酒和驾车有关）以及心脏病、中风等危急病例，而且还需要解决医院周围社区出现的一些问题，如呼吸障碍、割伤、撞伤和其他小外伤病人，他们占了每天就诊病人的主要部分。除了这些病人以外，其他就诊者就是居住在医院附近以及请不起私人医生的居民。这些人通常来看感冒、醉酒，甚至解决夫妻间的问题。这些变化使医院成为现代医疗保健体系的中心。如今的医院是具备高科技技术和复杂医疗模式的特殊机构，这种变化是不可避免的。科技的发展创造出越来越多的尖端仪器设备，即使较大的私人诊所都没有财力和人力购买和维护这些仪器设备，只有医院才能提供这类服务。科学技术创新的稳步发展不仅使医生、护士和技术人员等各个医疗卫生职业越来越专业化，而且还迫使先进技术装备都集中在现代化医院里。

近年来，科学技术的发展更是给医学界带来了雷轰般的巨大冲击。其中，人类基因组计划可能是 20 世纪 90 年代最卓越的科技成就。一方面，在这项成就的实现过程中，几种工程产品的开发起了关键的作用，包括自动测序仪、自动取样器以及数据库和序列拼接软件等。另一方面，基因组计划的成功使得生物医学工程人员开始关注细胞和分子水平的研究，而不仅仅局限于器官和系统水平的研

究，从而引起了又一个重大转变，展现出新的前景。比如，现在可以根据每个人的 DNA 来设计不同的药物治疗方案（见图 1.7）。另外，纳米技术、组织工程以及人造器官的发展也清楚地表明各种科学幻想将不断成为现实。但是，要有效地利用这些新技术，首先必须全面了解这些大量涌现的信息和创新将给社会和经济带来怎样的后果。

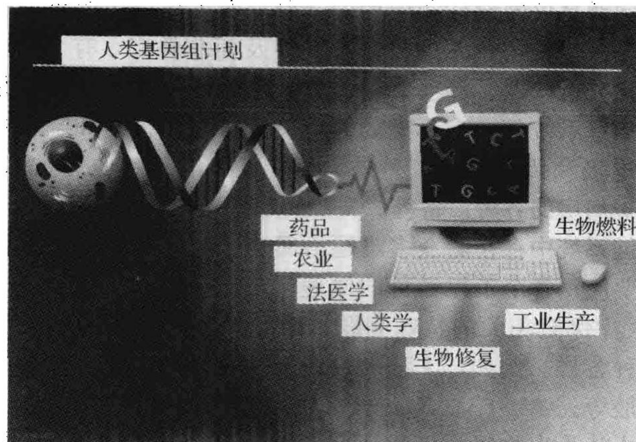


图 1.7　人类基因组计划的应用前景
（摘自 http：//labmed. hallym. ac. kr/genome/genome-photo/98-1453. jpg）

如果你展望一下，就可以看到科学技术的发展可能对医疗卫生行业产生的巨大影响（见图 1.8）。例如，通过与地区医疗中心相连的闭路电视通信网络，远在乡村的病人同样可以获得高水平的医疗服务。多项普检技术的发展为大众提供了预防医学，可以使那些真正需要使用医院设施进行诊断和治疗的病人住进医院。病人个人信息和护理记录的自动化管理为医生提供了病人状况的详情，包括住院情况和家中的情况。建立中央病历记录系统之后，即使病人变换了住址，或是离家途中因病就医，医生都能够方便快捷地获得病人的病历。另外，组织工程近年来也受到极大的关注，预计在 21 世纪的前 25 年里该领域会以指数形式发展。它是生物学、化学和工程技术的结合，通过生物材料、细胞、生长因子的单独使用或者联合使用来修补和恢复活体组织，或者使组织再生。上述例子只是医疗保健领域可能达到的、高质量的、成本合理的众多潜在

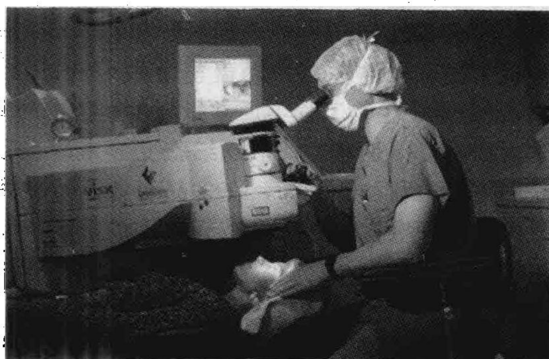


图 1.8　激光手术——医生的新工具（摘自 http：//riggottphoto. com/corporate/lgimg6. html）

技术发展中的一小部分。[⊖]

1.3 什么是生物医学工程学

对于生物医学工程技术人员而言，医疗保健行业如今所面临的各种问题都极为重要，因为这些问题是仪器设计和应用的基础，而工程应用的核心就是要开发各种仪器设备。这种与医学紧密相关的仪器开发涉及面非常广，有大型的复杂设备，如临床化验的自动化分析仪器、多项普检设施（也就是可以检测许多化验项目的试验中心）以及医院信息系统等；有小型的简单装置，如在科研和临床中用于监测特定生理信号的记录电极和传感器等；另外，还有复杂的远程监测技术，以及急救车、手术室和重症监护病房的各种设备等。

美国医疗保健系统中存在的众多难题给属于工程职业的生物医学工程师们提出了挑战。他们需要应用各个工程学科的原理、知识和方法（比如电子、机械和化工等），来解决与医疗保健相关的特殊问题，因此，工程师和医护人员的合作机会很多，并且形式各种各样。例如，工程师可以开发一种新的医学成像设备，或者给残疾人设计人工假肢等。

虽然，很多人认为生物医学工程研究领域所包含的范围比较清楚，但是，也有不少人持相反观点，他们认为生物医学工程学的有些领域与其定义不相符。例如，在生物工程高等教育指南中定义了下列名称：生物医学工程学（biomedical engineering）、生物工程学（bioengineering）、生物学工程（biological engineering）以及临床工程（clinical engineer）（即医学工程（medical engineer））。虽然 Pacela 在这个指南中把生物工程学定义为涵盖整个大范围的广义名称（参见本章后面的推荐阅读资料），但是，生物工程学通常指与生物技术和基因工程密切相关的偏向于基础研究的学科。其目标是通过动植物细胞变异来改良动植物，或者培养出新的微生物，以便获取更大的利益。例如，在食品行业，改良发酵用的酵母菌；在农业界，通过用有机生物处理植物的方法来减轻霜冻的损害，从而增加农作物的产量等。显然，生物工程人员的工作对提高人类的生活质量具有重要的作用。这个专业的前景难以彻底看清楚，其主要目标有如下几个：

- 开发动植物的改良品种，用于食品生产；
- 发明新的用于医学疾病诊断的试验方法；
- 通过克隆细胞生产人工疫苗；
- 通过生物环境工程保障人类和动植物的生命安全，避免有毒物质的污染；
- 蛋白质表面交互作用的研究；
- 酵母和杂交细胞生长的动力学分析和建模；
- 固定化酶技术的研究；

⊖ 有关生物医学工程学发展历史上主要事件的详细介绍，请参考 Nebekar 2002 年的综述论文。——原文注

• 医用蛋白和单克隆抗体的开发。

生物医学工程学这个名称的综合性似乎最强。生物医学工程师的职业目标是，应用电学、化学、光学、机械和其他工程学原理来认识、改造或者控制生物系统（也就是人或动物）。在医院和医疗中心工作的生物医学工程师被称为临床工程师更恰当，不过，现实中并不一定遵循这种理论上的区分，如今在美国医院工作的这类专家很多仍然被称为生物医学工程师。

生物医学工程师的研究领域极其广泛，虽然他们在 20 世纪五六十年代主要从事医学仪器的开发，但是，如图 1.9 所示，现在他们已经扩展了许多新兴领域，包括

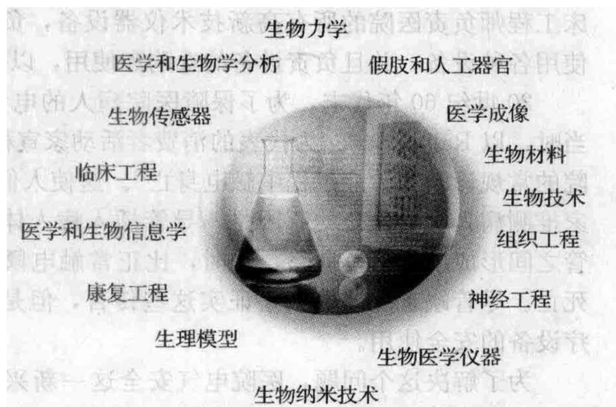


图 1.9　生物医学工程学所涉及的领域

• 工程系统分析的应用（即生理建模、仿真以及生物控制等）；

• 生理信号的检测和监视（也就是生物传感器和生物医学仪器）；

• 利用生物电信号的处理技术提取诊断信息；

• 开发治疗和康复的技术和装置（即康复工程）；

• 开发用以代替或者增强人体功能的装置（即人工器官）；

• 病人数据的计算机辅助分析和临床诊断（也就是医学信息学和人工智能）；

• 医学成像，即解剖结构和生理功能的图像显示；

• 新的生物制品的开发（也就是生物技术和组织工程学）。

生物医学工程师开展的研究工作有：

• 研究用于植入式人工器官的新材料；
• 开发用于血液分析的新型诊断仪器；
• 开发分析医学数据的软件；
• 为保证安全和疗效，进行医疗器械安全性能的分析；
• 开发新的诊断成像系统；
• 设计用于病人监护的遥测系统；
• 设计生物医学传感器；
• 开发用于疾病诊断和治疗的专家系统；
• 设计施药的闭环控制系统；
• 人体生理系统的仿真建模；
• 运动医学的仪器设计；
• 新型牙科材料的开发；
• 设计用于残疾人交流的辅助设备；

• 肺流体动力学的研究；
• 人体生物力学的研究；
• 人造皮肤材料的开发。

以上所列并没有完全包括生物医学工程涉及的所有研究工作，生物医学工程师可以发挥才干和技能的领域还有很多，实际上，他们的工作与所处的环境密切相关。在医院和医疗中心工作的生物医学工程师，也就是临床工程师尤其如此。现在的临床工程师负责医院的所有高新技术仪器设备，负责培训医护人员，使他们能够安全使用各种设备，并且负责设备的选购和使用，以便提供安全有效的医疗卫生服务。

20 世纪 60 年代末，为了保障医院病人的电气安全，工程师开始进入临床环境。当时，以 Ralph Nader 为代表的消费者活动家宣称："每年至少有 1200 名美国人在医院的常规诊断和治疗过程中触电身亡"。这使人们对电气安全的担忧达到了顶峰。大家推测病人触电致死的原因是，导管插入病人体内之后，就在体外到邻近心脏的血管之间形成了低阻抗通路，这时，比正常触电阈值低得多的电压就可能使病人触电死亡。尽管缺乏统计数据来证实这些传言，但是，这种呼吁使医务人员开始重视医疗设备的安全使用。

为了解决这个问题，医院电气安全这一新兴产业几乎在一夜之间就形成了。美国国家消防协会（National Fire Protection Association，NFPA）等机构制定了有关医院电气安全的标准。电气安全分析仪生产厂家和设备安全顾问等，都热切地希望为医院的各种医疗安全提供服务。有些公司立即开发了新产品用来保障病人的安全，特别是专用于配电系统的装置，比如隔离变压器等。为了缓解人们的恐惧情绪，医疗卫生机构认证联合委员会（Joint Commission on the Accreditation of Healthcare Organizations）（当时称为医院认证联合委员会）将美国国家消防协会制定的规则作为电气安全的标准，进一步规定医院至少每 6 个月检查一次所有用于病人身上或邻近病人的仪器设备的电气安全。为了满足这项新要求，医院管理人员考虑了几种可供选择的方案：雇用医疗仪器厂家来完成这些电气安全的检查工作；与公共服务机构签订合同，让其提供服务；由医院内部员工提供这些服务。大多数大型医院都选择了第 3 种方案，并且建立了独立的部门来提供电气安全所需的技术服务。

就这样，一个新兴工程学科——临床工程学诞生了，很多医院都建立了临床工程部。当这些部门开始运行时，人们很快就发现，电气安全故障其实只是医疗仪器所带来的许多问题中的一个。而且，当时医护人员对仪器缺乏全面的了解，也不懂得维护。简单的目测往往只能看到破损的旋钮、磨损的电线以及液体渗漏迹象等。许多设备并没有按照厂家的规定运行，也没有根据厂家的建议维护。总之，电气安全问题只是冰山一角。因此，临床工程人员的工作迅速发挥了作用，到了 20 世纪 70 年代中期，医疗设备使用前后的全面检查已经成为常规，并且形成了切合实际的检查程序。在此发展过程中，那些临床工程先驱们在医院里所起的作用越来越大。作为医院队伍的新成员，他们的工作包括：

• 积极参与开发符合成本效益的方法，发展医疗技术；

• 给医院管理人员提供咨询，根据医院的条件和特定的技术规格购买医疗设备；

• 启动科技新方法的应用，并且协助标准制定部门的工作；

• 培训医护人员，使他们能够安全有效地使用医疗仪器。

20 世纪七八十年代，在以下主要事件的推动下，临床工程学突飞猛进：

• 美国退役军人管理局（Veterans' Administration，VA）认识到临床工程师对于 VA 医院系统的正常运行至关重要，并将整个国家分为几个生物医学工程地区，每个地区安排一名生物医学工程主管工程师监管各个医院的所有工程技术问题。

• 在整个美国，绝大多数较大的医疗中心、医院和一些拥有 300 张以上病床的较小的医院都设立了临床工程部门。

• 医疗卫生职业人员，也就是医生和护士，在使用现有技术和引进新发明时急需临床工程师的援助。

• 建立了临床工程师的资格认证体系，确保临床工程师在任职期间一直能够胜任工作。

20 世纪 90 年代，临床工程学作为一个专业不断完善，在医学与生物工程国际联合会（International Federation of Medical and Biological Engineering，IFMBE）下设立了美国临床工程学协会（American College of Clinical Engineering，ACCE）和临床工程部（Clinical Engineering Division）。

如今临床工程师为临床医护人员提供了广泛的工程技术服务，并且已经成为整个医院非常重要的一部分（见图 1.10）。现代临床工程师既掌握医院内部具有的技

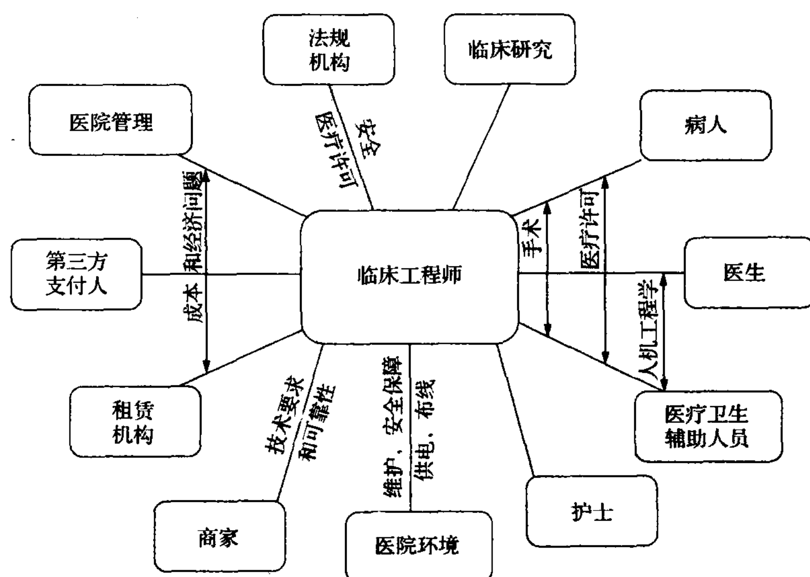


图 1.10　医疗环境中临床工程师的工作所涉及的范围

术资源，又了解医院之外厂家的技术资源，使医院可以有效地利用其绝大部分资源。

总之，生物医学工程学是工程学的一个交叉学科分支，它以工程学和生命科学为基础，从理论、研究、到创新应用，其工作包括了研究、开发、实现和运行。与医学职业一样，任何人都不可能成为整个生物医学工程领域的专家，只能成为某个方面的专家。因此，生物医学工程宽广的研究领域中涌现出了各种各样的专家。不过，由于该学科的交叉性质，这些专家的兴趣和工作内容都有很多相互关联和重合的地方。例如，致力于生物传感器开发的生物医学工程师可能会与开发假肢的工程师联合，以便研究检测和利用生物电信号驱动假肢的方法。再比如，致力于临床化学检验自动化的工程师也许会与那些开发专家系统的工程师合作，以便设计专家系统，辅助临床医生根据检验结果作出诊断。这种例子举不胜举。

生物医学工程师还有一个更重要的作用，就是发现医疗保健系统中目前还存在的那些可以用现有的工程技术解决的问题或需求。提供低费用高质量的医疗保健服务，这一直是医学界的奋斗目标，生物医学工程职业的出现给人们带来了希望。通过正确引导，如果生物医学工程师能够面向预防医学、急救服务等方面的问题，那么，他们就可以为进一步提高医疗保健系统的工作效率提供更好的方法和技术。

1.4 生物医学工程师的职能

广义上，生物医学工程专业主要培养三类人才：医疗保健系统的临床工程师；工业界的生物医学设计工程师；研究型科学家。不过，目前也按照生物医学工程师所履行的三种不同职能进行分类。第一类，也是最常见的一类工程师，可以称为“解题人”。临床工程师和生物医学设计工程师最可能成为这类人，他们与生命科学家之间保持着传统的服务关系，生命科学家发现并提出问题，由工程师应用特定的知识和技术来解决。但是，如果要使“解题”成功，双方必须具有很好的信息交流和沟通。生物医学工程师必须了解生物方面的情况，作出正确的判断，发挥自己的才能，解决所给的问题，并且用生命科学家可以理解的语言阐述所设计的解决方法。如果不能做到这些，那么“生物医学”这个冠名就名不副实了。

第二类人员比较少，可以称为“技术企业家”，工业界的生物医学设计工程师最可能成为这类人。他们认为，生命科学家和医生所接受的工业技术方面的教育与现有的技术潜能之间存在着相当大的差距，生命科学家不太可能结合现有的技术水平来提出问题。因此，科技企业家通过考察生物学和医学某些前沿领域的问题，自己寻找和确定先进技术可以发挥作用的地方，自己提出问题，然后解决问题。先设计，再用硬件和软件来实现。最后，这些企业家还必须证明他们的研究成果是有用的，说服医学界接纳这些成果。与上述“解题人”的境况不同，企业家的工作是有风险的，他们的成果没有现成的消费者。但是，冒险一旦成功，那么不论是科学上还是

经济上，都会前进一大步，比按照传统程序的进展速度要快得多。科技企业家的工作性质决定了他们应该掌握大量工程学和医学方面的知识，并且具有丰富的医疗仪器方面的工作经验。

第三类生物医学工程师称为“工程科学家”，在科研机构和大公司研究所最多，他们主要致力于应用工程学原理和技术来研究生物学过程。他们很喜欢用的一种方法，就是为所研究的生物系统建立一个合适的物理模型或者数学模型。然后，利用这个模型，通过仿真和计算，揭示生物系统的特性，这些特性如果用解析计算或者直观推导的方法来分析就会过于复杂。仿真的结果可以指导他们的实际生物系统的实验设计；反过来，实验的结果又可以用来改进仿真模型。如此反复，就可以一步步深入揭示生物学机制。

如果实验工作很繁重，难以施行，或者具有危险性；那么，数学模型还可以预测各种变化对于生物系统的影响。因此，工程科学家收获的是对于生物系统的进一步认识。并且，数学表述是一种紧凑简洁的语言，容易被其他人理解。由于要完成生物实验常常需要开发尖端测量技术，工程科学家的工作不可避免地会与仪器开发相关联。工程科学家必须与生物学打交道，这是最基本的。如果他们的研究成果最终将用于临床，那么更是如此。如果只是片面强调数学分析的优美，而忽视研究工作的临床应用意义，那是不够的。这类生物医学工程师是生物科学家真正的伙伴，他们已经成为很多科研机构研究队伍中不可缺少的成员，旨在开发新的实验和技术，用于揭示人体的生物学秘密。

上述生物医学工程师的每一种角色虽然对于生物学知识的掌握程度各有不同的要求，但是，每位工程师都必须是掌握重要工程技术专业知识的熟练专家。因此，在培养这个领域不同层次的新专业人员时，既要使学生学到足够多的生物学知识，同时又不能牺牲工程基础技术的学习，这使生物医学工程专业的教学计划制定经常会遭遇难题。我们正在迈入一个新的历史时期，面临着迅速增长的老龄人口，不断提升的社会和经济期望，以及疾病的预防、诊断和治疗方面更多技术开发的迫切需求。在这种形势下，生物医学工程师的培养和任职必不可少。这不仅因为这些人可以增进人类对于生命系统的了解，而且因为他们形成了一种充满希望的载体，可以加速知识向应用技术的转化。

与医生和护士一样，生物医学工程师最终的职责是为社会服务。这是一种职业，而不仅仅是熟练的技术服务。为了有效地发挥这类新生人员的作用，医学界的医护人员和管理人员应该充分认识这个新职业的重要性及其培训目标。生物医学工程学具有的巨大潜力、挑战和前景给我们带来的不仅是技术收益，而且还有人道主义效益。

1.5　生物医学工程的专业状况

生物医学工程师是专业人员。专业人员的定义是，经过相同强化培训、具有同种价值和技能的群体。判断一个人是不是专业人员，主要看他是否具有某种给定专

业的价值。专业人员必须具备专业价值，并且凭借其技术才能获得相关的许可证书。专业人员通常遵循与其工作相关的各种科学标准，专注于其技术能力范围内的工作，不受情绪影响，尊重客观现实，并且将客户利益放在自身利益之上。

从事医学技术设计、开发和管理的人员大致包括3类：科学家、商人和专业人员。首先看一下科学家与专业人员之间的区别。科学家追求的是知识，其价值体现在提供证据并与同行进行交流。而专业人员则是给客户提供服务，解决客户自己无法处理的问题。虽然科学家和专业人员都要运用知识、技能和专长。但是，科学家是凭才能工作，并向同行报道研究成果；而律师、医生和工程师等专业人员是为外行的客户服务。为了避免由于外行人员缺乏特定的知识而引起的麻烦，从事各种专业的人员通常由正式的国家许可机构来管理。专业人员和科学家都必须说服对象接受他们的发现。专业人员支持并遵循特定的道德规范为社会服务；而科学家则通过游说说服同行接受他们的观点。

以医学职业为例，其专业人员的主要职责是治病救人。这些人不仅有责任创建、发展并执行有关职业规范，而且还应该在一定条件下不计个人得失为公众服务。为了确保服务质量，医学界非常重视许可制度和认证程序。也许有人认为，这是医学从业人员自己在执行社会监控机制，但是这并不意味着社会的其他方面对医护人员的工作没有监督和控制作用。

专业人员的最后一个特征是职业道德。医生总是尽量宽容和支持病人，因而经常会在病人的愿望和社会利益之间陷入矛盾之中。例如，如何做到既尊重临终病人的意愿又不促进病人的死亡，这是医护人员不得不面对的道德问题。第2章将仔细讲述医疗技术带来的这类伦理道德问题。

按照时间顺序，我们可以根据以下6个关键事物的建立情况来确定某个专业的发展状况：第一所培训学校；第一所大学；第一个地方专业协会；第一个国家专业协会；第一部国家许可法规；第一部正式的伦理道德准则。

在这种发展过程中，最初的培训学校和大学一般注重基础知识学习的重要性。通过创新和改革，学校和教师会逐渐将理论知识与实践结合起来。以后，那些推进专业培训的人又会建立专业协会。协会规定职业机构的任务，也就是提高新成员的素质，重新定位成员的职责，让那些技术上稍有欠缺的人员去干比较常规的、较容易的工作；并且，平息内部和外部争端。其中，内部争端可以是那些坚持原有制度的人与那些追求变革和创新的人之间的矛盾。于是，此时就出现了许可和认证等机制。人们相信这样的机制可以确保专业的最低标准，提高专业的地位，并且保护外行人。

建立专业的最后一项工作就是设定正式的道德准则，一般包括的条例有淘汰不称职或无道德的从业人员的条例，减少内部竞争的条例，以及保障客户权益、强调为社会提供理想服务的条例。道德法规通常在专业成形的最后阶段制订。

生物医学工程专业的发展已经经历了上述6个关键步骤。20世纪50年代末，该专业最初由一群主要对医学电子学感兴趣的人员发起，现在已经从零星的群体发展为完善的组织机构。全世界大约有48个国际性协会为日益扩增的生物医学工程师队

伍服务。如今，生物医学工程专业所涉及的领域非常广泛，近年来，组织工程学和人工智能等许多以前不属于该专业的新学科，也已经成为该专业的一部分。

在其发展过程中，专业协会起到了主要的作用，它将该专业不同学科的成员聚集起来，分享在探索新技术应用时所获得的知识和经验，为改善人类的健康状况和提高生活质量而努力。通过生物医学工程协会、美国医学与生物工程学会以及隶属于 IEEE 的生物医学工程学会等专业机构的工作，国内和国际水平上的各个协会之间的合作如今越来越活跃。

1.6　生物医学工程专业学（协）会

1.6.1　美国医学与生物工程学会

美国拥有世界上最大的生物医学工程师队伍，为这支队伍服务的侧重点各不相同的主要专业机构有：美国临床工程学会（American College of Clinical Engineering），美国化学工程师协会（American Institute of Chemical Engineers），美国医学信息学协会（American Medical Informatics Association），美国农业工程师协会（American Society of Agricultural Engineers），美国人工器官学会（American Society for Artificial Internal Organs），美国机械工程师协会（American Society of Mechanical Engineers），医疗仪器促进协会（Association for the Advancement of Medical Instrumentation），生物医学工程学会（Biomedical Engineering Society），IEEE 生物医学工程学会（IEEE Engineering in Medicine and Biology Society），康复与辅助技术发展交叉学科联合会（interdisciplinary Association for the Advancement of Rehabilitation and Assistive Technologies），生物材料学会（Society for Biomaterials）。为了将这些不同学（协）会中的各种生物医学工程内容统一起来，人们在 1992 年创建了美国医学与生物工程学会（American Institute for Medical and Biological Engineering ，AIMBE）。AIMBE 的主要目标是包罗和统一美国的生物工程队伍，处理公共政策问题，为促进人类健康和提高生活质量不断开拓工程应用途径。AIMBE 的联系方式是：1901 Pennsylvania Avenue N. W. , Suite 401，Washington，D. C. 20006（http：// aimbe. org/；Email：info@aimbe. org）。

1.6.2　IEEE 生物医学工程学会

美国电气电子工程师学会（Institute of Electrical and Electronic Engineers，IEEE）是世界上最大的国际性专业机构，它一共包括 37 个协会和理事会，其中 IEEE 生物医学工程学会（IEEE Engineering in Medicine and Biology Society，EMBS）是最重要的国际组织之一，为全球 8000 多位生物医学工程会员提供服务。EMBS 主要致力于物理学和工程学的原理和方法在生物学和医学中的应用。该学会每年资助召开一次重大的国际学术会议，同时也在世界各地协助主办针对不同主题的地区性

学术会议。学会出版的主要学术刊物有：生物医学工程学汇刊（Transactions on Biomedical Engineering）（月刊）、神经系统与康复工程汇刊（Transactions on Neural Systems and Rehabilitation Engineering）（季刊）、生物医学信息技术汇刊（Transactions on Information Technology in Biomedicine）（季刊）、纳米生物学汇刊（Transactions on Nanobioscience）（季刊）、IEEE 医学与生物工程杂志（IEEE Engineering in Medicine and Biology Magazine）（双月刊）。还有与其他学会联合出版的学术刊物，包括医学成像汇刊（Transactions on Medical Imaging）、神经网络汇刊（Transactions on Neural Networks）、模式分析与机器智能汇刊（Transactions on Pattern Analysis and Machine Intelligence）。IEEE EMBS 行政办公室的联系方式是：IEEE，445 Hoes Lane，Piscataway，NJ，08855 - 1331 USA（http：//www.embs.org/；Email：emb-exec@ieee.org）。

1.6.3 生物医学工程协会

为了使代表生物医学利益和工程学利益的两个方面在协会组织中具有同等的地位，美国生物医学工程协会（Biomedical Engineering Society，BMES）于 1968 年应运而生。正如 BMES 协会章程中所述，其首要目标是“促进生物医学工程学知识的增长和应用”。该协会每年春季和秋季定期召开两次学术会议。另外，每年还与 AIMBE 和 EMBS 等其他生物医学工程学会联合主办各种专题的学术会议。与 BMES 有关的主要学术刊物有：生物医学工程纪事（Annals of Biomedical Engineering），是刊登生物医学工程各个领域创新性研究论文的月刊；生物医学工程通报（BMES Bulletin），是广泛刊登生物医学工程研究、新闻和相关事件的季刊；还有 BMES 会员目录（BMES Membership Directory），是罗列协会成员个人联络信息的年刊。BMES 的联系方式是：BMES，8401 Corporate Drive，Suite 225，Landover，MD 20785 - 2224，USA（http：// www.bmes.org/default.asp；Email：info@bmes.org）。

上述各个生物医学工程学（协）会的工作对于生物医学工程师职业地位的不断提升具有非常重要的作用。因此，鼓励所有生物医学工程师，包括该专业的学生，都积极成为这些学（协）会的会员，并投入到真正的职业工作中来。

习　　题

1. 请说出下列各个历史时期发明的一项特殊生物医学工程技术，并说明它们的基本工作原理，论述其对于医疗保健系统所产生的影响。

(a) 1900～1939；(b) 1945～1970；(c) 1970～1980；(d) 1980～2003。

2. 请综述计算机技术对于医疗保健系统所产生的影响，举例说明计算机在该领域的应用及其发明时间。

3. “基因工程”这个名称意味着它具有工程作用，对吗？这方面的工作应该属于生物医学工程学领域吗？

4. 请详细论述人类基因组计划对医学新技术的发展已经产生的作用以及预期将会产生的作用。

5. 请你展望一下，工程学和生命科学中哪些方面的进展将对临床医疗和生物医学研究产生最大的影响?

6. 医院的组织结构有 3 个主体：董事会、管理人员、医务人员。请说明 3 者各自的主要职责。临床工程部门应该属于哪个主体? 请给出解释。

7. 根据定义，临床工程师应该具备哪些特征?

8. 请列举至少 7 项临床工程师特有的工作。

9. 生物医学工程师可分成 3 类：解题人、技术企业家和工程科学家。请分别举例说明他们的工作成就。

10. 下列人员是否符合专业人员的定义，为什么? (a) 注册护士；(b) 生物医学技术人员；(c) 呼吸治疗师；(d) 医院管理人员。

11. 请列出从事生物医学工程工作必须掌握的知识，在哪些教学环节中可以获得这些知识? 怎样才能较好地学习管理技能?

12. 请提供某个生物医学工程专业协会主页的材料，并列出该协会下一年的主要活动事项。

13. 你认为生物医学工程师在未来的医疗卫生系统中将会承担怎样的职能?

14. 请论述由于经济条件限制所造成的医疗保健方面的得失。

15. 请分析医学技术的经济成本因素和经济效益。

推荐阅读资料

Aston, C. (2001). Biological warfare canaries. *IEEE Spectrum* **38:10**, 35–40.

Bankman, I.N. (2000). *Handbook of Medical Imaging*. CRC Press, Boca Raton, FL.

Bronzino, J.D. (2005). *Biomedical Engineering Handbook,* 2nd Ed. CRC Press, Boca Raton, FL.

Bronzino, J.D. (1992). *Management of Medical Technology: A Primer for Clinical Engineering*. Butterworth, Stoneham, MA.

Carson, E. and Cobelli, C. (2001). *Modeling Methodology for Physiology and Medicine.* Academic Press, San Diego, CA.

Laurenchin, C.T. (2003). Repair and restore with tissue engineering. *EMBS Magazine* **22:5,** 16–17.

Nebekar, F. (2002). Golden accomplishments in biomedical engineering. *EMBS Magazine* **21:3**, 17–48.

Pacela, A. (1990). *Bioengineering Education Directory.* Quest Publishing, Brea, CA.

Palsson, B.O. and Bhatia, S.N. (2004). *Tissue Engineering*. Prentice Hall, Englewood, NJ.

The *EMBS Magazine* published by the Institute of Electrical and Electronic Engineers, edited by John Enderle, especially Writing the book on BME, **21:3**, 2002.

第 2 章　道德问题与伦理问题

Joseph Bronzino 博士，专业工程师

本章目录

本章主要内容：

- 道德和伦理的定义与区别；
- 功利主义和康德主义这两大主要哲学流派的基本思想；
- 医生、护士和生物医学工程人员的伦理准则；
- 现代社会的一些道德难题，包括死亡的重新定义、晚期病人的护理、人体试验等，它们都源于行善和不伤害这两条基本道德准则；
- 目前新型医疗器械开发与使用的管理政策所涉及的道德问题。

医疗技术的飞速发展给人类带来了崭新的医学时代。例如，随着材料科学的发

展，我们已经能够制造人工假肢、人工心脏瓣膜和人工血管等产品，从而产生了所谓的“备用件”手术治疗法；无数病人的常规疾病诊断都用上了各种各样高新技术成像设备；随着呼吸机、心脏起搏器和人工肾脏等先进辅助和治疗设备的发展，许多病人的生命得以延续。

不过，科学技术发展带给人类的并非全都是福音，它们也带来了重要的道德问题。如今，人类已经能够开发心血管辅助装置或者实施器官移植来维持晚期病人的呼吸和心跳，这迫使人们不得不重新审视死亡、生存质量以及尽力与怜悯等问题的含义，并且开始考虑病人拒绝治疗的权利（即预立遗嘱）和参与临床试验的权利（即知情同意）。可见，科技进步使人类医疗保健所涉及的道德问题变得越来越复杂，这给医务人员、生物医学工程师，甚至整个社会提出了新的令人烦恼的道德难题。

本章的目的是分析某些与新型医疗技术应用相关的道德问题，我们并不提供解决问题的办法和建议，只是表明每一种科技进步的结果都会影响人类价值观的核心问题。

科学技术与伦理学具有密切的联系，它们是人类文明社会的两大成就。科学技术是人类非凡才能的体现，它与人类伦理道德分别属于两个范畴，两者相互对峙，并不相交。一方面，通过伦理学分析，可以看到科学技术的潜力和限制；另一方面，科学技术以伦理道德为指导来实现人类的目标和利益。

在科学技术与伦理学的这种相互对峙过程中，我们希望从事医疗新技术开发和负责病人护理的人们能够通过批判地审视所谓的“公认”观点，来获得自己的结论。本章将首先介绍有关道德和伦理的定义，然后较详细地论述对于生物医学工程人员特别重要的一些道德问题。

2.1 道德和伦理的定义

自古以来，人们就一直在考虑生命的本质及其意义。德国哲学家康德（Immanuel Kant，1724—1804）曾经就此提出 4 个基本问题：我能认识什么？我应该做什么？我能希望什么？人是什么？部落是最早的人类社会组织形式，部落的风俗习惯中所包含的复杂的行为准则体现了早期人类社会中就存在上述这些问题。大约在公元前 600 年，希腊人把有关这些问题的众多原始思想、态度和观点简化成为一些很好的理论，并将它们整合成为普遍适用的智慧体系，这就是哲学。作为航海家和殖民者，希腊人与众多不同民族和文化具有密切的接触，在他们所接触的各种各样社会习俗、法规和制度的冲击下，希腊人开始审视并比较这些社会中的所有人类行为，而且把由此形成的那部分哲学思想称为伦理学。

伦理学的英文名词“ethics”来源于希腊文“ethos”，它是“习俗”的意思。另一方面，拉丁文中“习俗”一词是“cos”，其复数形式是“mores”，它与希腊文的“ethos”意思等同，是英文中道德的“moral”和“morality”的词根。虽然伦理和道德这两个名词常被互用，但两者之间也存在区别。

哲学家将伦理学定义为一门特殊的学问，而将道德作为该学问的主题。由那些

持久不变的人类相互关系所产生的习俗称为道德。例如，说真话、欠债必还、尊敬父母、尊重他人的权利和财产等，都是我们所生活的社会的道德规范。大部分社会成员通常都认为这些行为规范不仅是习俗，而且是正确的。因此，道德是指人们认为正确的、好的、应该遵循的习俗。

我们中的大多数人都能够遵循这些行为准则，根据基本道德规范调整自己的生活方式。许多人甚至宁愿牺牲自己的生命也不愿意背叛这些道德规范，他们不仅将道德规范作为自己的行为准则，同时也以此衡量他人的行为。那些漠视公认的行为准则的人则被认为是行为不轨者，在很多情况下这些人会因为社会不能接受的行为而受到惩罚。例如，行使了社会定义的“犯罪行为”的人被称为罪犯，通常要受到严惩。不过，这些行为准则的判定并不是一成不变的，它们必须经常调整，以适应社会的变化，从而避免它们阻碍社会的改革和发展。

道德观代表了社会的行为准则，而伦理学是研究人类行为的对与错、善与恶的一门学问。伦理学不关心人类行为的具体评判准则，只是客观分析人们“应该做什么”。由此可见，伦理学其实代表了道德的哲学观，因此，常被称为道德哲学。

考虑如下 3 个问题：严重畸形的婴儿是否应该存活下去？是否可以终止晚期病人的治疗而让他们死去？是否可以进行人体试验？这些问题属于道德问题还是伦理问题呢？根据上述定义可知，这 3 个问题都属于道德评判问题。

哲学家们认为所有的道德评判都可以规范化，可以直接用善、恶、对、错等特征价值指标来表示。例如，以下就是一些常用的规范化评判：

- 偷东西是错的；
- 每个人都享有接受教育的权利；
- 自愿安乐死不应该合法化。

每一个评判都表明了一种价值观，也就是表达了人们支持或者反对某种事件的态度，因此，它们具有指导行为的作用。

不过，要做出道德评判，必须了解人类社会的正确道德标准。那么，什么是正确的道德标准呢？为了回答这个问题，我们先来了解一下目前主宰伦理学的两大对立学说，即功利主义与康德主义。功利主义注重行为的后果，是一种结果论；而康德主义则注重行为的动机，是一种非结果论。结果论认为可以达到最佳结果的行为总是正确的道德行为，其主要论点就是，没有特定的行为可以自动归为道德的或者不道德的，某种行为的对与错完全由其效果决定。

根据功利主义学说，在任何情况下，可以根据以下两步来决定应该采取的行动：首先，确定可以选择的行动；其次，判断哪种行动的结果能够获得最多快乐、最少痛苦。能够给“最多的人带来最大好处”的行为就是正确的道德行为。医疗仪器设计、开发和使用的主要动机就是提高医疗保健水平，因此，根据功利主义的观点，评价医疗技术的标准显然就是大家所公认的对健康有益，也就是能够维护人类的幸福感。

由此可见，功利主义者提倡用结果验证方法。只要治疗方法能够达到最佳结果，就应该使用。然而，很多人并不认为结果一定能够验证方法，他们认为无论结果怎

样好，人的权利是始终不可侵犯的。

与功利主义对立的伦理学学说就是非结果论，其支持者否认道德价值完全由行为的结果决定的观点，他们认为道德价值还与其他因素相关，不同意凡是能够获得最好结果的行为就是正确道德行为的观点。该伦理学学说主要基于康德主义，它坚持认为在道德观上人类具有其特别珍贵的东西。根据康德理论，人具有其他动物所没有的权利。例如，人不能为了食物而互相残杀和食用，不能为了体育运动而互相恶性竞争，不能为了医学而进行人体试验。这些都是基于人权观点的道德判断。总而言之，人拥有特殊的不可侵犯的权利，只因为他们是人。

伦理学的主要任务就是确定为了尊重人的独特尊严所必须采取的行为。对于大多数康德主义者而言，这决定了人在相处时必须遵守的制约，换言之，就是明确每个人的基本道德权利。

可见，上述两种伦理学思想提供了一些道德评判的基本准则。但是，如果没有明确的道德判断，我们在处理问题时常常会面临困境。医学上，在判断病人治疗和保健方法的对与错时经常会遇到道德难题。很多情况下，医务人员必须在两种治疗方案中做出选择，而这两种方案似乎都不能达到令人满意的结果。例如，是延续生命重要，还是阻止痛苦重要？如果治疗可能会缩短寿命，那么，停止治疗是否正确？如果不治疗可能导致死亡，那么病人是否仍然有权利拒绝治疗？根据我们现有的判定对与错的准则，这些问题似乎都没有明确的答案。难怪康德会提出这样的难题：“我应该做什么？”

案例 1：干细胞研究

如图 2.1 所示，精子与卵子相结合之后，就会发生受精过程并启动胚胎发育。精子一进入卵细胞，立即就会引起细胞膜离子通道的开放，从而使细胞膜去极化，

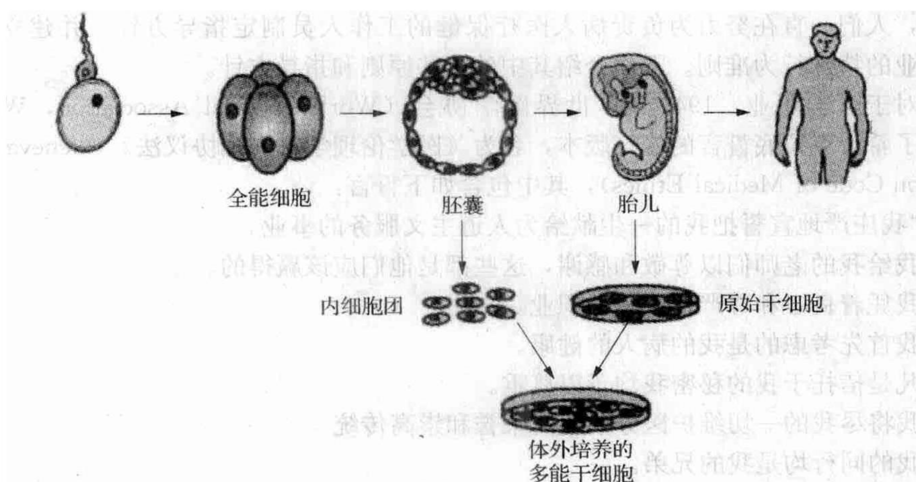


图 2.1 利用内细胞团产生多能干细胞的示意图

（摘自网站：http：//www.nih.gov/news/stemcell/primer.htm）

阻止其他精子再进入卵细胞。然后，卵细胞的 DNA 开始复制，并在大约 36h 之后发生第一次细胞分裂。以后，发育过程继续进行，细胞开始卵裂，也就是细胞重复发生分裂，并且，其分裂周期只包含 DNA 合成的 S 期和有丝分裂的 M 期，基本上跳过一般细胞分裂所包含的 G1 期和 G2 期。因此，细胞的总体积并没有增加，只是分裂成越来越小的子细胞，这些细胞被称为分裂球（blastomere）。

受精 5 天之后，胚胎的细胞数目达到数百个，这些细胞紧密连接成上皮组织的形式，排列在一个腔体周围，构成所谓的囊胚。囊胚中存在一团突向腔内的细胞，这些细胞称为内细胞团，它将发育成为胚胎。外层细胞则形成滋养层，最终会变成胎盘。囊胚的内细胞团可以被分离出来，在体外进行培养，并由此获得胚胎干细胞。

必须注意，如果细胞分裂不断继续下去，细胞就会发生定型和分化。细胞发生分化后表现出其命运注定的特异性细胞功能。干细胞的定型与分化相关，但两者有所不同。已经定型的细胞群移植之后，不会被同化成为其他细胞，而是会生长成为其命运注定的器官组织细胞。

由于采集胚胎干细胞时会破坏胚胎，于是，就出现了以下问题：

• 胚胎是否属于活人，它是否具有其他年龄阶段活人所享有的人权？这个问题可以分别根据功利主义和康德哲学两种观点来分析。

• 是否任何对人类幸福感有利的研究都应该去追求？

• 美国联邦政府是否应该支持这类研究工作？

医学中的各种道德难题当然不是现在才出现的，它们贯穿着整个医学史。长期以来，人们一直在努力为负责病人医疗保健的工作人员制定指导方针，并建立了医学职业的特殊行为准则。下面介绍其中的一些原则和指导方针。

对于医学职业，1949 年，世界医学协会（World Medical Association，WMA）通过了希波克拉底誓言的修订版本，名为《医学伦理学日内瓦协议法》（Geneva Convention Code of Medical Ethics），其中包含如下誓言：

“我庄严地宣誓把我的一生献给为人道主义服务的事业。

我给我的老师们以尊敬和感谢，这些都是他们应该赢得的。

我凭着良心和尊严行使我的职业。

我首先考虑的是我的病人的健康。

凡是信托于我的秘密我均予以尊重。

我将尽我的一切维护医务职业的荣誉和崇高传统。

我的同行均是我的兄弟。

在我的职责和我的病人之间不允许把对宗教、国籍、种族、政党和社会党派的考虑掺杂进去。

即使受到威胁，我也将以最大的努力尊重从胎儿开始的人的生命。

我绝不利用我的医学知识违背人道法规。

我庄严地、自主地、并以我的名誉做出上述保证。”

美国医学协会（American Medical Association，AMA）1980 年通过了一套医学伦理学准则，并在 2001 年 6 月进行了修订。下面是修订前后两套准则的对比。

修订之前的准则

1980 年美国医学协会代表议会通过

医学职业早就签署有伦理学声明，它们主要根据病人的利益建立。医师作为医学职业的成员，不仅必须对病人负责，而且必须对社会、对其他医务人员、同时对自己负责。下列是美国医学协会通过的准则，它们并不是法律条文，只是受人尊敬的医师的基本行为标准。

1）医师应该抱着同情心和对人类尊严的敬重全心全意地提供称职的医疗服务。

2）医师应该诚实对待病人和同事，并努力揭发人品不端和不称职的医师以及他们的欺骗行为。

3）医师在尊重法律的同时，也有责任对那些违反病人根本利益的要求寻求变更。

4）医师应该尊重病人、同事和其他医务人员的权利，并且在法律限定的范围内保守病人的秘密。

5）医师应该不断学习和运用科学知识，并促进其发展。与病人、同事和公众交流相关信息，积极获取咨询，并利用其他医务人员的才能。

6）除了急救情况以外，医师在所提供的合适的病人护理条件下，可以自由选择服务对象，选择合作人员和实施医疗服务的场所。

7）医师有责任参与旨在改善社区状况的各项活动。

修订后的准则

2001 年 6 月 17 日美国医学协会代表议会通过

医学职业早就签署有伦理学声明，它们主要根据病人的利益建立。医师作为医学职业的成员，必须首先对病人负责[⊖]，也必须对社会、对其他医务人员、同时对自己负责。下列是美国医学协会通过的原则，它们并不是法律条文，只是值得尊敬的医师的基本行为标准。

1）医师应该抱着同情心和对人类尊严和权利的敬重全心全意地提供称职的医疗保健服务。

2）医师应该拥护职业道德标准，诚实对待所有职业关系，并努力向相关机构报告人品不端和不称职的医师以及他们的欺骗行为。

⊖ 下画线表示修订的内容。——译者注

3）医师在尊重法律的同时，也有责任对那些违反病人根本利益的要求寻求变更。

4）医师应该尊重病人、同事和其他医务人员的权利，并且在法律限定的范围内保守病人的秘密和隐私。

5）医师应该不断学习和运用科学知识，并促进其发展。承担医学教育的义务，与病人、同事和公众交流相关信息，积极获取咨询，并利用其他医务人员的才能。

6）除了急救情况以外，医师在所提供的合适的病人护理条件下，可以自由选择服务对象，选择合作人员和实施医疗保健服务的场所。

7）医师有责任参与旨在改善社区状况和提高公共卫生水平的各项活动。

8）在治疗病人时，医师应该把对病人的责任放在首位。

9）医师应该支持任何人都有权享受医疗保健服务的观点。

对于护理职业，美国护士协会（American Nurses Association，ANA）1976年正式通过了护士章程，其中的声明和解释给护士履行护理职责时的行为和关系提供了指导。详见下文：

“序言：护士章程建立在对人的本性、护理、健康和社会的认识的基础之上。护理服务的接受者和给予者都是拥有基本权利和义务的个人和团体，他们的价值和境况在任何时候都必须得到尊重。护理的目的是促进健康的恢复、预防疾病、缓解痛苦。本章程的声明及其解释给护士履行护理职责时的行为和关系提供了指导，护理职责包括职业的道德义务和护理质量。

1）护士应以尊重人的尊严的态度为病人服务，不论病人的社会和经济地位如何，个人品性和疾病的性质如何，对病人都应该一视同仁。

2）护士应该能够判断病人信息的私密性，保护病人的私人权利。

3）无论何人如果要以不称职的、不道德的或者非法的行为危害他人的健康和安全时，护士都应当采取措施保护病人和公众。

4）护士有义务和责任承担病人护理的决策并实施工作。

5）护士应该能够胜任护理工作。

6）护士应该具备个人的能力和资历，能够作出正确的评判，寻求咨询、接受任务，并能够指派他人做护理工作。

7）护士应该参加那些有利于本职业知识发展的活动。

8）护士应该努力提高护理水平。

9）护士应该努力创造和维护有利于提高护理质量的工作条件。

10）护士应该努力防止公众误报和误传信息，维护护士职业的高尚。

11）护士应该与其他医务人员和群众合作，为满足公众的健康需求，共同努力促进社区和国家的发展。”

这些准则作为护士行为的指导方针，表明了为人类服务和尊重人的生命的原则。读了这些行为准则后，很难想象谁还能对这些有关护理病人的主要目标的总结性陈

述做出什么改进。然而，有些人认为这些准则并没有给目前医务人员所面临的很多道德难题提供答案。很多时候，护士不可能同时履行所有应尽的基本职责。例如，如果有一位脑部遭受严重损伤的病人正在用人工设备维持生命，而此时其他病人也需要使用这台设备，那么，根据上述指导方针，如何完成保护生命促进健康的护士职责呢？可能有人会争辩说，是否给予病人治疗是医生决定的，而不是护士决定的，但是，这两种职业的工作人员都与病人的护理密切相关，他们都关心这种决定的最终含义。

随着对于职业行为伦理道德重要性的认识日益加深，生物医学工程人员也建立了各种职业道德准则。这些准则一般都由数条规则组成，它们既表明了所有职业人员应该遵守的最低标准，也提出了他们应该努力的目标，为从事该职业的人员提供了道德行为的指导方针。例如，如下是美国临床工程学会签署的道德准则：

"作为美国临床工程学会的成员，我同意遵守以下道德准则，我将：

1）正确体现我的职责、权力、经验、知识和受教育水平；

2）努力防止科学技术的应用给人类造成的风险；

3）揭露那些可能影响信息的提供和获取的各种利益冲突；

4）尊重信息的私密性；

5）致力于人类医疗保健的发展；

6）致力于改善医疗技术的管理和设施建设，从而控制其成本；

7）促进临床工程职业的发展。"

这些准则有助于规范道德行为，但是，显然它们不可能提供针对每一种具体情况的道德指导。一个旨在使其成员道德行为规范化的职业，不能将其成员对于道德的认识只局限在职业准则的范围内，它必须为成员们提供各种资源，使他们在具体的特殊情况下有能力决定需要遵守哪些准则，从而使他们在指导性准则含糊不清或者根本没有的情况下，能够做出符合伦理道德的判断。

2.2 两条基本道德原则——行善与不伤害

自古希腊文明的西方医学开始，在各种为卫生保健系统工作人员制定的准则和誓言中就一直包含了"行善"和"不伤害"这两条道德原则。行善就是提供帮助；不伤害就是避免伤害。这两条原则的起源可以追溯到古典名作《希波克拉底文集》(Hippocratic Corpus)。该文集虽然以公认的西方医学创始人希波克拉底命名，不过，医学历史学家其实并不能肯定文集中的所有内容（甚至包括希波克拉底宣言）究竟是否出自其本人之手。可以断定文集中的部分内容写于公元前 6 世纪，但其他内容却被认为写于公元初期。历史学家认为文集中的很多具体道德准则并不是古希腊和古罗马时期一般医师的实际行医方法，也不是他们的道德思想。

无论如何，至少有一部分古代医师将如下这句名言作为基本医学道德准则，这就是"面对病人，要习惯于做两件事：一是救人，二是至少不伤害。"随着希腊文明

的衰败和基督教的兴起，行善和不伤害逐渐成为医疗行为的基本道德原则。虽然这两条原则在古希腊和古罗马医疗技术中只是附属品，然而，基督教所强调的人类同情心和兄弟情使这两条原则逐渐变成了惟一可接受的行医动机。甚至在如今制定的西方医务人员行为准则中，仍然与以前行医者的誓言和准则中一样，强调行善和不伤害。

传统医疗保健伦理学中对于不伤害的强调程度明显超过行善，这与医学发展历史有关。历史上，医学的伤害能力远远超过其保护和恢复健康的能力。行医者的很多治疗方法对病人显然有极大的危险，并且似乎不会带来什么好处，真正有效的治疗方法实在太少。在这种情况下，当然要尽可能强调不伤害，优先考虑不伤害，而不是行善。

现代科学技术的发展彻底改变了这种状况。实验研究获得的知识、临床试验得到的结果以及统计方法验证的数据等都在不断充实医学知识，指导医疗活动。这种医学与科学之间的持续合作已经成为大量医疗技术发展的主要源泉，遍及整个医疗保健系统，使得医学的治疗、预防和康复能力显著提高，从而也将“行善”推到了医学道德的前沿。有人甚至提出，“不伤害高于一切”的老观点已过时，应该用“病人应该获得最好的治疗”的新观点来代替。不过，日新月异的医疗技术也带来了很大的不确定性，人们很难断定究竟什么会给病人带来最大的益处、最小的伤害。换言之，随着医学“行善”能力的增强，医学技术也带来了究竟什么是有益且无害治疗的许多争议。

以上概述了医学伦理和道德的一些基本原则，下面我们介绍医疗技术的应用所带来的几大具体道德问题。

2.3 死亡的新定义

虽然医学上很早就开始观察并鉴定死亡，但是，许多医生并不了解生命起始和终结的哲学观点。由于医学是一门经验性的临床科学，过去，医务人员在医学上似乎没有必要考虑死亡的基本概念，只要知道死亡的事实就足够了。以前，活与死被看做两种极端情况，两者之间存在着无限隔断的鸿沟。如今，医务人员利用先进的医疗技术可以延长病人的生命，活与死的定义也因此发生了变化。

毫无疑问，医疗技术的应用已经从死神手中挽救了无数病人的生命。这一点从人均预期寿命（average life expectancy）的变化趋势上就可以得到证实。例如，如今美国的人均预期寿命男性为74.3岁，女性为76岁；而在1900年，男性和女性的人均预期寿命都只有47岁。发达国家的科学技术已经成为社会文化的一部分，这使得婴儿的死亡率也大大下降。医疗技术所提供的人工环境使早产儿不再受到死亡的威胁。如今的医学技术不仅可以减少早期死亡，还可以有效地推迟那些不可避免的死亡的发生时间。例如，心脏起搏器、人工肾脏等各种各样医疗仪器使无数病人的生命延长了很多年。医学技术的发展所取得的成功甚至可以使医务人员能够长时间延续濒死病人的“生命体征”。不过，这些医疗过程也产生了有关病人生存质量的重要

哲学问题。

假如有位病人在一次交通事故中脑部严重受伤，赶到事故现场的急救人员发现病人已经昏迷，但心脏仍然在跳动。病人被送到医院急诊室之后，值班医生的检查结果表明病人的心跳和呼吸生命体征稳定，CT 扫描显示大范围严重颅脑损伤，头皮电极测得的脑电图也表现出明显的异常。在这种情况下，显然会出现这样的问题：病人究竟处于什么状态？是死还是活？

再举一个开胸手术的例子。假设手术时病人身上连接了人工心肺机，外科医生正在修复病人受损的心脏瓣膜。在复杂而漫长的手术期间，原来正常的脑电图监测波形突然几乎变成了一条直线，显示出非常微弱的脑电活动。但此时人工心肺机仍然维持着病人的所谓生命体征。这时医生应该怎样办？仍然认为病人还活着并继续手术，还是认为病人已经死亡？

临床上，上述这些情况越来越多，从而促使医务人员重新审视死亡的定义。实际上是拖延死亡的先进医疗技术提出了死亡的重定义要求。这没什么可奇怪的，其实在人类历史上，死亡的定义一直与医学知识和医疗技术的发展密切相关。以前，长达数百年时间里，人们一直认为没有呼吸就是死亡。后来，人体生理学知识不断得到积累，而且医疗技术的发展可以使失去呼吸的人重新苏醒过来，于是，人们逐渐将心脏停止跳动作为确定死亡的主要指标。然而，随着支持性疗法、心血管辅助装置和器官移植等医学技术的进一步发展，死亡的定义又发生了变化。

随着人类对于人体结构及其功能的不断深入了解，现在大家都知道血液的主要成分之一是氧气，任何人体器官如果持续缺氧超过一定的时间就会停止工作并死亡。具备高级功能的脑组织特别容易受到缺氧的伤害，供血中缺氧时间超过很短的 3min，脑组织都会产生不可逆的损伤。于是，人们又将死亡的征兆从心脏转移到脑。自从医学上能够利用脑电图监测脑的活动之后，死亡的定义中就引入了一个新的指标。脑死亡观念的支持者认为人脑是生命真正不可缺少的东西。一旦脑受到不可逆的损伤，那么，脑所控制的功能，以及记忆、感觉、思维和认知等与人性相关的功能就都会受损。

这样，临床上被广泛接受的死亡的定义就变成了肺、心脏和脑的自发活动全部丧失。也就是在宣布任何人死亡之前，其心、肺、脑 3 个主要器官都必须已经不可逆地停止工作了。虽然肝和肾等其他器官系统的受损可能会通过其对心、肺、脑重要功能的致命影响最终导致病人的死亡，但是，临床死亡的定义不包括这些内容。

然而，随着先进呼吸机技术的发展，遭受不可逆脑损伤的病人依靠呼吸机几乎可以一直维持下去，这种病例如今越来越多。于是，医疗技术的新发展再一次要求我们重新审视死亡的定义。

1968 年，哈佛医学院特别委员会发表了一份报告，极大地推进了重新定义死亡的工作。该报告提出了另一种基于脑功能的死亡定义，被公认为是医学技术发展史上有关死亡问题的一个里程碑。

简而言之，该委员会制定的死亡标准包括下列内容：病人必须无感知并无反应，

也就是处于不可逆昏迷状态；撤除机械呼吸机时，病人必须没有呼吸活动；病人必须没有任何反射反应；病人的脑电信号变为平坦的时间必须至少超过24h，表明其没有脑电活动。只有当这些条件都满足时，才能宣布病人死亡。

同时，该委员会还强烈建议，参与死者后续的器官或者组织移植工作的医生不能是宣布病人死亡并撤除呼吸机的人。这样，供体病人的死亡就不会受到器官移植动机的逼迫。这一点很重要，必须将护理受者病人的权利和责任与护理供者病人的权利和责任彻底分开。

脑死亡概念的出现表明，仅仅具有生物活性的人并不能称为活人，这其实说明植物人的生活并不是人的生活。换言之，肢体完整且存活着的脑死亡病人并不属于活人。我们很多人都理所当然地认为真正活在世上的人必须具有“完整的大脑功能”，正是这种思想使得死亡的定义逐渐从以心跳和呼吸为指标，转变到面向脑的指标。

实际上，完全且不可逆的脑功能丧失被称为脑干死亡（brainstem death），也就是全脑死亡，或者简称脑死亡，它已经被公认为死亡的法律标准。按照这个标准，脑死亡的病人在法律上与死尸没有什么区别。即使医疗技术的干预可以维持其呼吸和血液循环功能，这种人在法律上可以与死尸一样处理。很多人认为这个法律标准是符合道德原则的，一旦发生脑干受损，人脑就根本不能工作了，除非使用人工辅助设备维持生命状态，否则机体的各种调节机制都会失去。因此，利用机械方法维持脑死亡病人的状态只不过是推迟不可避免的最终结果的发生时间而已，对于其人性、人格和意识的丧失都毫无作用。是机械干预使这样的人体看起来与死尸不一样，但是，机械通气状态下的死尸终究还是死尸。

虽然已经形成共识，脑死亡就是死亡，处于脑死亡状态的人其实就是一具死尸，但是，棘手的病例仍然存在。例如，某病人处于持续性植物状态，也就是所谓的新皮质死亡（neocortical death）。这种病人虽然遭受了严重的脑损伤，但其脑功能仍然足以维持呼吸和循环系统的功能，并不需要机械辅助。然而处于持续性植物状态的病人对于外界刺激不能产生有目的的响应，也没有自我意识。不过，病人的眼睛可能会间歇性地睁开，并且可能表现出睡眠觉醒周期。某些病人甚至可以自发地产生打哈欠、咀嚼和吞咽等动作。与毫无反应的脑干死亡病人不同，这种病人可以诱发出各种简单或复杂的反应。但是，这样的病人重新恢复意识的机会非常渺茫。在人工进食、肾透析等方法的维持下，这种病人的状态可以维持数十年。著名的Karen Ann Quinlan案例就是这种情况。

1975年4月，年轻女子Karen遭受了严重的脑损伤，并且进入一种慢性植物人状态，她不再具有任何认知能力。她父母接受了医生的诊断，认为她没有恢复的希望了，于是，就寻求法律的许可，要求新泽西州立医院重症监护室的医护人员撤除她赖以生存的呼吸机。

初审法庭以及随后的新泽西高级法庭都同意撤除Karen的呼吸机，于是呼吸机就被去掉了。但是，医院里负责照顾她的护士反对这个决定，抢先一步，事先就帮

助 Karen 戒掉了对呼吸机的依赖，这样，当呼吸机撤除时她已经可以自主呼吸。因此，当时 Karen 并没有死，她又活了 10 年，直到 1985 年 6 月，最终死于急性肺炎。这一次没有给她使用治疗肺炎的抗生素。

如果脑干死亡是死亡，那么，新皮质死亡也是死亡吗？同样，这个问题并没有明确的答案。这里也需要确定这种病人与死尸之间存在哪些不同之处，从而证明在道德上是否允许像死尸一样处理。根据传统上呼吸和循环不可逆停止工作的死亡标准，处于持续性植物状态的病人并不是死尸，道德上不能像死尸一样处理。另一方面，根据脑干死亡标准，处于新皮质死亡的病人也不是死尸。因为该标准的关键是脑损伤的严重程度必须达到身体各种调节机制的丧失。

那么，与利用机械辅助设备维持呼吸和血液循环的脑死亡病人相比，新皮质死亡的病人缺少了哪些特征呢？这是一个社会必须回答的问题。有了答案之后，才能清楚怎样的治疗对于新皮质死亡病人是属于行善和不伤害。

2.4　晚期病人与安乐死

如今，晚期病人经常会发现自己处于一种奇怪的境地，一方面，社会制度和医学技术在帮助他们对抗死亡，而另一方面，拥有先进科学技术的现代化医疗系统却可能造成病人及其家庭相当大的经济负担以及心理和生理上的痛苦。而且，为了延续生命，先进的医疗技术经常可以拖延临终病人的死亡过程，使病人更加有损尊严。这种情况又引起了一个医学道德难题，如果医务人员停止晚期病人的治疗，或者给病人注射致死剂量的药物，这究竟是对还是错呢？

这个问题已经成为如今社会需要考虑的主要问题之一。虽然交通事故、用药过量、酗酒、谋杀、自杀等造成的死亡随处可见，但是，对于大多数人而言，一般都是寿终正寝，或者死于某些慢性病。当生命的旅程即将终结时，人们希望那一天到来时所有的烦恼都会结束，满怀同情心的家属们也常抱有这种愿望。但是，抱着惟一目标的医学治疗却经常会粉碎这种愿望，无论如何，医学治疗都要尽可能延续生命。于是，很多希望尊严死的病人的要求经常与如今的医疗标准不相容。

例如，下面这个案例就是如今病人、家庭、医务人员和社会经常会共同面临的一个问题。假如有一位中年男子动脉破裂引起脑溢血，并失去了知觉，而且一直住院治疗，再也不能够恢复知觉，处于新皮质死亡状态，也就是慢性植物人状态。塑料袋中的营养液通过手术放置的胃管直接滴入其胃部，以维持他的生命。病人的护理每天需要 7.5h，护理工作包括剃须、清洁口腔、梳理、清除肠道和膀胱的排泄物等。再假设他妻子通过法律途径要求停止其丈夫的所有治疗，包括停止营养液和水的供给，使其丈夫的机体可以彻底死亡。她向法庭陈述了有力的证据，表明在这种情况下她丈夫所希望的就是这样的结果。

这个案例产生的主要道德问题就是病人生存质量的下降是否已经达到道德上允许的有意结束生命的程度。清醒的时候，病人曾经清楚地告诉家人和朋友，他宁愿

死也不愿意在无法恢复意识的状态下通过机械装置维持生命。那么，是否可以按照他的意愿行事呢？首先必须明确怎样的生存能力和质量值得维持下去，反过来，这些生存质量的缺失是否足以作为结束生命的依据，至少这可能是病人自己的意愿。如果这个问题没有明确的答案，那么，传统的医学伦理学准则，也就是行善和不伤害，就没有指导意义，因为无法确定终止维持病人生命的支持对病人究竟是有益还是有害。

对于多数人而言，争取生命是正确的职业观，他们认为医学始终应该致力于应用先进技术尽可能延长人的生命。不过，对于晚期病人，这并不是惟一的办法。如果要保护个人权利，那么，医生尽可能延长生命的某些做法所引起的一些道德问题就必须解决。只要各种药物、手术、呼吸机、人工肾脏、心脏起搏器等的使用能够延长病人的生命，就尽可能实施，这不应该是医务人员的惟一目标。医务人员应该给予每位病人合理的生存质量。正是出于这种新观点，安乐死再一次成为医学界争议的问题。

“安乐死”的英文“euthanasia”源于两个希腊文字“美好”和“死亡”。其实很多原始社会都不同程度地施行安乐死，例如，古希腊在 Cos 岛上每年要举行一次老人和病人的聚餐，让他们饮用一种毒药。甚至亚里斯多德也主张对严重畸形儿实施安乐死。其他古代社会也用类似的方法对待老人，当某些人不再有用时就会被抛弃。然而，随着基督教在西方社会的传播，人们对安乐死产生了一种新的态度。因为圣经中耶稣基督的信条是“不能杀人”，“杀人者应判死刑”，于是安乐死的实施大为减少。

如今社会，安乐死更多的是意味着“尊严死”。当生命只能依靠机器维持，似乎不再有价值时，才能实施安乐死。在许多情况下，安乐死意味着终止生命是为了避免病人在不治之症的晚期忍受不必要的痛苦，因此它具有解脱痛苦的含义。

在讨论安乐死的道德问题时经常将其分成主动安乐死和被动安乐死两类，两者的区别在于实施安乐死的行为是积极的还是消极的。如果通过注射针剂等方法加速病人的死亡，相当于杀死病人，这是积极安乐死。如果不采取可以有效推延病人死亡的措施，任病人自行死亡，这是消极安乐死。这两种安乐死的区别在于主动和被动。在被动安乐死中，医生并没有做任何导致病人死亡的事情，死亡是由病人所患的疾病引起的。相反，在主动安乐死中，医生做了某件事，从而导致了病人的死亡。例如，医生给癌症病人注射致命的针剂后引起病人死亡，这就是医生导致病人死亡。而如果医生只是终止治疗，那么，死因就是癌症。

主动安乐死中，必须有人做了某件导致病人死亡的事；而被动安乐死中，病人的死亡由疾病引起，而不是人的行为引起的。这种说法对吗？假如某位医生有意不治疗一位晚期病人，然后病人死亡了。再假设，该医生为自己辩解说“我没做任何事，病人是病死的，我没有使他死亡。”根据目前的法律和道德标准，这样的辩解毫无说服力。这位医生肯定会像主动杀死了病人那样受到谴责。因此，医生采取措施一直治疗到病人死亡为止，这才是被人理解的行为。

安乐死还可以分为自愿与非自愿两类。如果为病人实施的安乐死只是为病人好，但不是病人自己的意愿，那么，这种安乐死就是非自愿的。可见，从道德方面看，非自愿安乐死与不道德的杀人没有什么区别。但是，当病人已经没有能力表达同意与不同意时该怎么办呢？假如某位晚期病人丧失了意识，无法表达自己的意愿。那么，是否允许加速他的死亡呢？如果有充分的证据能够表明病人事先曾经同意这种做法，那么，这是可以的。例如，病人可能告诉过亲戚和朋友，在某些情况下不应该继续延长他的生命，甚至可能有遗嘱，或者有录音带、录像带等记录资料。如果有这种事先同意的证据，那么，实施安乐死就不违背道德。这种情况就属于自愿安乐死。

病人在签署遗嘱时必须具有清醒的意识，并且在签署遗嘱与疾病发作之间的这段时间里没有改变主意，只有这样遗嘱才有效。此外，遗嘱的证明人不能是病人死亡的利益获得者。正如遗嘱本身所述，它并不是具有法律约束力的文件，它只是被动的请求，只有道德上的说服力。不过，遗嘱的倡议者们相信，在决定让病人自行死亡时，遗嘱可以减轻医务人员和家人的负罪感。

那些赞成安乐死的人认为个人权利和选择自由是最重要的，他们把安乐死看做结束病人痛苦的一种仁慈的方法，尊严死比不断遭受痛苦而逐渐衰弱直至死亡更有吸引力。这些倡议者们把每个人看做拥有独特思想和个性的理性人，他们坚持晚期病人应该有权控制自己生命的结束。

另一方面，那些反对安乐死的人强调谁也无权结束他人的生命。

有些人则根据已建立的准则行事，他们提醒那些对病人承担责任的人必须尽力挽救生命。他们的意见是医务人员不可能在认为安乐死是合理的同时仍然履行自己的誓言。如果晚期病人还活着，那么，至少还有机会寻找可能有用的治疗方法。反对安乐死的人觉得安乐死的合法化会破坏医生与病人之间的信任。如果病人不相信医生和护士会尽全力给予治疗，并且知道，一旦病情恶化，医务人员就会失去信心，并可能决定还是让病人死亡好，那么，病人会怎么想呢？那些反对者还质问安乐死究竟是真正对遭受痛苦的病人有利，还是只是减轻病人家属痛苦的一种办法。他们相信，无论破坏多么微弱的生命，如果只是为了解除他人的精神痛苦，实在是不公平的事情。

还有很多人担心，如果安乐死合法化，那么，很难定义和建立执行安乐死的明确的基本条例。而且，一旦社会接受了任何形式的安乐死，就会引出许多其他问题。即使只允许被动安乐死，它也可能被用于各种收治智障病人和老人的国立医院和机构，那些地方现在住着成千上万没有生活希望的人。另外，被动安乐死的合法化可能会使更多的父母寻求方法，结束身患不治之症的孩子的生命，从而摆脱烦恼；也会有更多的子女寻求方法，提早结束年老的或者身患绝症的父母的痛苦。例如，在纳粹德国，安乐死最初只是用于结束晚期病人的痛苦；但是，后来，安乐死的实施普及开来，以至于那些稍有不正常的人，如精神病人、犹太人等少数民族，都遭到杀害。显然，这种情况发人深省。

遗嘱样本

我的家人、医生、牧师、律师们：

如果将来某个时候，我没有能力再决定自己的未来，那么，就让这份声明作为我的意愿。如果我没有任何合理的希望能够恢复身体或精神上的残疾，我要求让我自行死亡，不要用大量的人工措施维持我的生命。死亡与出生、成长、成熟和衰老一样，是必然要发生的事情。我宁愿死，也不愿意因为病情恶化、失去自立以及绝望的痛苦而失去尊严。即使药物会加速死亡，我也请求给我使用能够减轻我最终痛苦的药物。本请求经过慎重考虑。虽然这份遗嘱没有法律约束力，但我希望关心我的人能够从道德上满足我的要求。我知道这会让你们承担沉重的责任，我希望这份遗嘱能够分担这种责任并减轻它给你们带来的内疚感。

签名________________

日期________________

证明人______________

2.5 医疗决策权

医疗决策的选择可能会极其困难，它们往往涉及令人不愉快的问题，而且，往往在我们的身心都最脆弱的时候出现。这些选择几乎总是涉及新的医学信息，使人感觉很陌生且无法抗拒。以下步骤有助于人们进行医疗决策：

1）获取所有数据，也就是明了各种医疗数据；

2）了解所有选择及其后果；

3）根据你自己的价值观，给每一种选择设定一个价值。

这 3 个步骤构成的“数据－选择－价值”途径关注的只是“如何”决策，而同样重要的是“谁”作出决策，必须有人对每一种医疗方案作出决策。理论上，作出决策的人应该是与其最密切相关的人，也就是病人。然而，多数情况下，却是其他人作决定，如配偶、家人、医生、或者是代表病人的其他人。这里，必须清楚医疗决定的同意权由 4 个层次组成：

1）最中心的第一层是病人；

2）第二层是利用病人先前的指示，也就是根据遗嘱等文件进行决策；

3）第三层就是他人为病人作出决定，也就是从自我控制变换为他人代理控制；

4）最后一层是法院和政府。当病人丧失能力，也没有明确的预先指示，该由谁作出决定不明确时，社会法令告诉我们法院和政府是我们最终可以求助的地方。

上述 3 个步骤和 4 个层次只是在混乱的医疗决策过程中加上了一些条理，它们有助于人们掌握控制权。

2.6 人体试验

在现代社会中，医学研究长期以来一直拥有崇高的地位，从预防脊髓灰质炎（即小儿麻痹症）的 Salk 疫苗和 Sabin 疫苗的成功开发，到人工器官的研制，医学研究的卓越成就受到公众的热情赞颂。不过，为了确定研究结果的有效性以及临床应用价值，新药和新型医疗仪器最终都必须经过人体试验。这里存在的问题，一是人是否应该参与旨在对自己或者他人有利的临床试验，二是要明确这些试验在什么条件下才能允许进行。

例如，假设有一位 50 岁女性病人患有严重的冠心病，那么，在对她可能有益也可能无益的新药或者新仪器试验中应该遵循什么指导方针呢？是不是只有那些看起来对她可能有益的试验才能做呢？为了评价某些诊断方法和仪器设备的有效性，即使它们对受试病人没有直接的益处，并且还有其他更容易接受的治疗方法可供选择，在这种情况下，可以在病人身上进行临床试验吗？

我们再考虑一个有关人体胎儿医学研究的问题。随着美国社会堕胎的合法化和胚胎研究技术的发展，这类研究已成为可能。那么，在什么情况下才能进行这方面的医学研究呢？例如，为了确定某些药物对胎儿的影响，是否可以让正在计划堕胎的妇女服用这些存在潜在危险的药物呢？流产的胎儿是否可以用于任何实验研究呢？

这些问题很难回答，但是，对病人的幸福负有责任的临床研究人员在试验新仪器和新方法时必须面对这些道德问题，同时也必须保护病人的个人权利。

案例 2：新生儿重症监护病房

新生儿出生体重不足往往由早产引起，长期以来，它是影响婴儿存活率的一个主要因素。体重不足的婴儿通常分成两类，一类是体重小于 1500g 的低出生体重儿（Low Birth Weight，LBW），另一类是体重小于 1000g 的极低出生体重儿（Very Low Birth Weight，VLBW）。护理这些婴儿时必须极其小心谨慎，要尽可能增加他们的存活几率。随着各种早产儿医疗护理技术的发展，如改进的体温调节技术和通气技术等，现在 LBW 和 VLBW 的死亡率已经大大降低。如图 2.2 所示，新生儿重症监护病房（Neonatal Intensive Care Unit，NICU）的建立也是一个重要的进步，它使得维持婴儿生命所需的设备都可以方便地集中在一起。



图 2.2　新生儿重症监护病房
（摘自 http：//www.hlarch.com/health1.htm）

NICU 中最重要的设备之一就是早产儿培养箱。这种装置一般用透明塑料制作，用于稳定婴儿的体温。其实，培养箱使医务人员不必用毯子包裹婴儿就可以保持婴儿的体温。培养箱还可以防止感染并保持生长环境的湿度。新生儿环境温度和湿度的稳定可以使婴儿身体保持一定的水分，避免水分的丢失。

很多新生儿都患有一种并发症，就是不能自己正常呼吸，要么完全不能呼吸，要么有呼吸暂停现象。所谓的呼吸暂停是一种非周期性的不规则呼吸模式。

容易发生呼吸暂停的婴儿必须受到密切监视，一旦他们的呼吸停止了，护士就必须冲到床边唤醒他们。但是，有时护士赶到现场需要数分钟时间。为了便于唤醒发生呼吸暂停的婴儿，生物医学工程技术人员开发了一种触觉振动器，它由呼吸暂停事件触发，会在婴儿脚底产生振动，从而自动唤醒婴儿。为了证明这种装置的有效性和安全性，必须进行人体试验。这时，需要解决如下问题：

谁负责这项试验研究？

这种试验的审批程序应该包括哪些内容？

知情同意的方针应该是什么？

在试验研究过程中如果装置有所改动，并且这些变化改变了装置在原来申报时的性质，那么，这些改动是否应该准许呢？

2.7 人体试验的定义和目的

究竟怎样的试验是人体试验呢？虽然试验方法有各种各样的变化，但是，一般认为只要在临床病人治疗过程中有意收集有关药物和器材使用效果的信息，就属于人体试验。以前，人体试验分成治疗性（therapeutic）和非治疗性（nontherapeutic）两类。治疗性试验是可能对病人直接有利的试验，而非治疗性试验则只是为了获取更多的知识，而不是直接有利于病人。两者的主要差别在于试验目的，而不是试验结果。

在整个医学发展史上，治疗性医学试验的例子层出不穷。例如，应用非常规放射治疗抑制恶性肿瘤，利用心脏起搏器提供正常心脏功能所需的电刺激信号，还有使用模拟肾脏功能的人工肾脏排除血液中的毒素等，这些医疗技术都曾经是对病人可能有价值的治疗新方法。经过临床试验，现在已经证实它们不仅对病人有益而且对整个人类社会都有利。

非治疗性试验也一直是推动医学发展的重要工具。例如，肝炎病毒感染、疟原虫感染以及心脏导管插入手术的效果等方面的实验研究都曾经对医学的发展产生过重要的作用，对于造福全人类的医疗技术的最终发展具有深刻的影响。

20 世纪 70 年代中期，美国“保护生物医学与行为学研究的受试人的国家委员会”（National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research）提出用“实践”（practice）和“研究”（research）这两个术语来

替代传统的治疗性和非治疗性试验。1986 年，Alexander Capron 引用委员会的报告内容写到：

"'实践'是指那些具有合理的成功期望，并且纯粹为提高病人或客户的幸福而进行的活动。在医学界，实践通常包括诊断、预防和治疗；在社会上，实践包括政府的转移支付和教育等项目。

相对而言，'研究'是指验证某个假设、获取结论，从而对基础知识的发展有所贡献的活动，基础知识可以表示为各种理论、原理、相互关系的描述等。由此可见，实践和研究处于两个极端，实践是应用已证实的科学技术造福于人类的活动，而研究则是为了增加知识而开展的工作。"

分成实践和研究两种活动的优点是可以使人们明确造福于病人的临床试验并不是只有治疗性活动，但这种分类法仍然是基于目的，而不顾结果。为了病人的利益而应用已证实的科学技术，这类活动是实践；为了增加理论知识而进行的活动则是研究；那么，不符合这两种性质的活动又是什么呢？

无效实践就属于这类活动，它包括预防和诊断治疗活动。使用"无效实践"一词的主要目的是为了保护病人的利益，向病人强调将要进行的试验还没有被证实是安全且有效的。对于那些将要接受无效实践的人，必须让他们适当了解情况，并取得他们的同意。

2.8　知情同意

长期以来，很多人一直认为知情同意是人体试验最重要的道德条件。为了使人体试验合法且合理，必须满足知情同意这个首要条件。所有成年人都具有同意医疗试验的法律权利，除非这种权利已通过法律程序被废除。因此，有关医疗试验法律权利的问题一般局限于未成年人。许多国家，即使不是所有国家，都有一些允许未成年人给予同意的有关条例。

知情同意的目的是保护受试人的权利，让他们有机会自己决定是否愿意参加任何医学试验。1964 年，世界医学协会在芬兰签署了一份有关人体试验的道德准则，旨在给这方面的工作提供指导方针。2000 年 10 月在苏格兰爱丁堡召开的第 52 界世界医学协会大会对这些指导方针进行了修订。

由于人体试验的结果对于推进科学知识的发展非常重要，因此，世界医学协会向全世界医生推荐了以下各小节介绍的基本指导方针。当然，必须注意，这些指导方针并不是指医生、科学家和工程师可以摆脱他们各自国家所规定的刑事责任、民事责任和道德责任。

案例 3：人工心脏

20 世纪 80 年代初，为了给当时最新研制的人工心脏 Jarvik-7（见图 2.3）挑选第一位候选人，有关部门专门成立了一个审查委员会。大家决定第一位接受人

工心脏移植的病人必须是已经濒临死亡的人，如果选择还可能活一年的人进行试验，那是不道德的，因为人工心脏很可能会杀死病人。

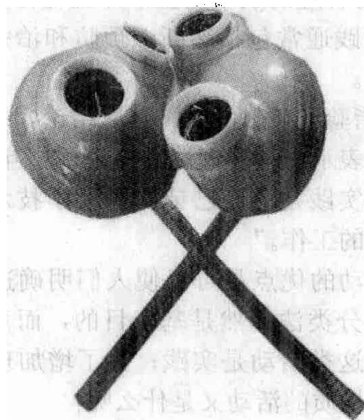


图 2.3 Jarvik-7 人工心脏

注：1985 年 8 月在美国 Arizona 大学的医疗中心，Jack G. Copeland 博士曾经将此人工心脏植入正在等待心脏移植手术的病人 Michael Drummond 身上。Jarvik-7 使 Drummond 一直活到一周之后获得捐献器官为止。经过数月手术应用之后，医疗中心将这个人工心脏交给了美国史密森学会，该学会将其视作人体“备用件”的成功典范（备用件是指替代人体器官功能的仪器设备），它也是医疗技术发展引起争议的一个例子（摘自网址 http：//www. smithsonianlegacies. si. edu/object description. cfm? ID=172）。

这是一种无效实践吗？

这种情况是否仍然需要知情同意？

手术一周后，病人 Barney Clark 从头到脚开始发生痉挛。每次痉挛发作期间，Clark 失去知觉的身体都会颤抖数小时。在接下来的几个月时间里，痉挛和精神错乱不断发生。最后，Clark 表示希望死去。虽然在一段采访录像中 Clark 确实肯定了人工心脏的作用，但他并不幸福，身上绑着一个大机器，几乎没有知觉，而且还要忍受痛苦。此后，1983 年 3 月 Barney Clark 死于多发性器官衰竭。

请利用本案例详细讨论“成功的标准”和生存质量的概念。

2.8.1 基本原则

1）涉及人体受试者的生物医学研究必须符合普遍公认的科学原理，必须已经经过充分的测试和动物实验，并且以全面的科技文献资料作为基础。

2）涉及人体受试者的每种试验过程的设计和实施方法都必须清楚地写成试验议定书，并递交给专门任命的独立委员会进行审核，以征求意见并获得指导。

3）涉及人体试验的生物医学研究必须由具备科研资历的人员承担，并且必须在具有临床资历的医务人员监督下进行。医务人员必须始终对受试者负责，即使受试

者已经签署了知情同意书，也绝不能把责任推给受试者。

4）除非试验的重要性与受试者需要承担的内在风险相当，否则人体生物医学试验研究就不能合法进行。

5）对于每个涉及人体试验的生物医学研究项目，都应该事先仔细评估其可能带来的风险，并将这些风险与受试者或他人可能获得的利益进行比较。必须始终将受试者的利益放在科学利益和社会利益之前。

6）必须始终尊重受试者保护自身不受损害的权利，应该谨慎采取预防措施以尊重受试者的隐私权，尽可能减少试验研究对受试者造成的生理、心理以及人格的影响。

7）医生应该回避参与人体试验，除非他们确信其中的危险性是可以预见的。而且，一旦发现试验的危险性超过其潜在的益处，医生应该终止试验研究。

8）发表受试者的试验研究结果时，医生有义务保证结果的准确性。与本宣言所制定的原则不相符的试验报告不应该发表。

9）任何人体试验中，每位受试者都必须被充分告知试验的目的、方法、预期的利益和潜在的危险，以及它可能带来的不适。并且，受试者还应该被告知有权拒绝参与研究，而且有权随时撤回同意书。然后，医生才可以取得受试者自愿给予的知情同意，最好是书面的同意书。

10）在获取受试者的知情同意时，医生应该特别注意他自己与受试者之间是否存在关系，受试者是否会迫于某些关系而签署知情同意，如果是这样，就应该由不参与研究工作并且与某些关系完全无关的医生去获取知情同意。

11）如果受试者无法律行为能力，那么，应该取得国家法律规定的法定监护人的知情同意。也就是当受试者在生理上或者心理上没有能力给予知情同意，或者受试者是未成年人时，根据国家法律的规定，应该从对受试者负责的亲属那里获得许可。

12）试验研究议定书必须包括有关伦理道德问题的声明，并且表明该试验遵守本宣言中阐明的各项准则。

2.8.2　结合治疗的临床医学研究

1）在病人的治疗过程中，如果医生判定某种新的诊断或治疗技术有希望挽救生命、恢复健康、减轻痛苦，那么，医生就应该有使用这种新方法的自由。

2）新方法可能带来的好处、风险和不适应该与现有最好的诊断和治疗方法的优点进行权衡。

3）任何医学研究都应该保证每位病人（包括对照组）能够获得已证实的最好的诊断和治疗方法。

4）如果病人拒绝参与研究，这绝对不能影响医生与病人之间的关系。

5）如果医生认为不能进行知情同意过程，那么，就应该在递交给独立委员会的试验议定书上写明特殊的理由。

6）只有确定了医学研究对于病人具有潜在的诊断和治疗价值之后，医生才可以将医学研究与病人的医疗过程结合起来，以获取新的医学知识。

2.8.3 非治疗性临床生物医学研究

1）在人体上进行纯科学性医学研究时，医生有责任保护受试者的生命和健康。

2）受试者必须是志愿者，可以是健康人，或者是病人，但试验的设计与病人所患疾病必须无关。

3）研究人员一旦确定继续试验可能会给受试者造成伤害，就应该立即终止试验。

4）在进行人体试验研究时，科学利益和社会利益绝不应该凌驾于受试者的利益之上。

这些准则可以归纳为人体试验的6条基本道德要求：人体试验必须以充分的实验研究和动物研究为基础，必须基于充分的科学数据，这样才能明确人体试验的研究重点，并且利用非人体方法获得尽可能多的结果；人体研究应该使用能够合理地提供所需信息的试验方法，如果所用方法不能提供所需信息，那就毫无意义，没有科学研究价值；研究工作必须由相关的科学技术专家操作；必须仔细评定该人体试验对于受试者、对于科学，以至对于社会的所有可预见的风险和利益，并且利益必须明显大于风险，而且不存在其他风险较小的方法能够获得这些可能的利益；受试者应该知情并自愿同意参加试验；如果受试者中途不愿意再继续试验，或者谨慎的研究人员认为继续试验可能会导致受试者受伤、残疾或者死亡，那么就应该立即停止试验。

遵守这些条件确实可能会限制医学发展的速度和广度，但是，我们的社会坚持执行这些条例是因为真正有价值的医学发展必须高度尊重人类的尊严。在这些条件中，获取受试者自愿的知情同意被公认为是最重要的保护措施之一。

根据这些准则的含义，有几类受试者可以自动被排除，他们不能参与医学研究项目。儿童以及思考能力受疾病影响的病人都可以被排除，因为他们不能准确理解试验内容。此外，那些与临床研究人员有关系的人，如研究人员的病人和学生等，也可以被排除。由于心理能力也包括受试者判断试验后果严重性的能力，因此，对于某些类型的医疗方法有些未成年人虽然有权利给予同意，但是，前提是他们必须能够理解其中的风险。

任何研究项目都必须明确定义其中的风险，病人必须能够毫无隐瞒地获得所有已知的信息。过去，很多情况下风险和利益的评估都由医务人员单独执行。风险和利益一旦确定之后，就以为病人会接受这些定义。如今，这种做法不再有效。尽管医务人员仍然必须评估他们所推荐方法的风险和利益，但是，具有最终决定权的是病人。当然，病人不可能决定这些医疗方法是否正确，作出这些评价所需的医学专业知识一般都超出了普通人的知识水平。但是，一旦某种医疗方法被推荐之后，病人就必须获得足够的信息来决定其中所期待的利益是否值得他们承担其中的风险。这样获得的知情同意才有效。

一旦获取并记录了病人自愿的知情同意之后，研究人员就有了如下保障：

1）知情同意代表了医学人体试验的法律许可。受试者以后不得攻击和殴打有关人员。

2）知情同意给予了有关人员在医疗和研究上使用试验数据的法律权利，受试者

等人以后不得提出侵犯隐私的起诉。

3）如果受试者没能从试验方法中获益，不得起诉。

4）在了解试验的风险并同意接受试验的情况下，如果受试者受伤，不得起诉。

5）在试验规则合理并已解释清楚的情况下，如果受试者由于没有遵守安全指示而导致受伤，那么研究人员可以免受起诉。

不过，医学研究的目标是否总能够与医生的传统道德义务相一致呢？研究人员和医生是否可能处于难以防守的境地呢？这里，受试者自愿的知情同意仍然是关键。因为，这种情况下，试验引起的后果都是他们自己选择的结果，不是他人强加的，实实在在是他们自己选择接受的。由于受试者的自主权受到尊重，他们并没有被迫成为他人利益的牺牲品。即使他们可能会为了他人的利益而受苦，这也是出于他们自己的意志，是他们自主行为的结果，而不是自主权受到限制或剥夺产生的结果。

为了保证知情同意的真实性，它必须是真正自愿的，而不是强迫的。并不是所有的强迫都像体罚那样明显，容易识别。受试者可能会担心没有别的治疗方法，或者担心不同意会疏远负责其治疗的医生，甚至会担心别人的谴责，这些都可能迫使受试者给出同意。这种胁迫经常难以察觉和纠正。

最后，受试者必须了解他们同意的究竟是什么。因此，他们必须根据所给予的充分的信息来作出明智的决定，决定是否参加试验研究。虽然不需要把研究人员所知道的所有信息都告诉受试者，但是，在不影响受试者作出决定的有效性的条件下，必须确定哪些信息应该提供给受试者，哪些可以省略。还有一个难题就是很难知道受试者是否有能力理解有关信息并在此基础上作出明智的选择。无论如何，任何情况下都必须努力保证提供充足的相关信息，并使受试者能够分析这些信息。这些事虽然很可能不能做到绝对准确和肯定，但是，必须全心全意作出努力，以防止人体试验触犯人类的尊严。

案例 4：秘密、隐私与同意

美国医疗保健行业近年来发生的一个巨大变化就是医疗信息系统中计算机应用技术的发展（见图 2.4）。医疗信息系统的计算机化确实提供了数据采集、存储和连接的强大工具，但是，这种系统的实现也存在挑战和风险。

例如，在某个中等城市，孩子们经常会到居民区附近医院的急诊室就医，过去这种就医方式存在的主要问题之一就是没有医疗记录，孩子后来在其他医生处看病时，医生就不知道以前的治疗情况。为了解决这个问题，有人提议建立儿童卫生保健网络，它可以使医生了解来自医院附近某个学校的儿童的医疗记录。该系统为每位在校儿童建立一个计算机化的医疗记录，当地医院的医生都可以查阅并更新这个记录。

请详细讨论这种系统应该注意病人的哪些隐私权问题。

请详细讨论应该在何处并如何处理同意权问题。



图 2.4 用于监视儿童卫生保健状况的先进技术

（摘自网址 http：//www.ustechlab.com/和 http：//www.bbc.co.uk/threecounties/do_that/2002/10/black_history_month.shtml）

2.9 医疗器械产品开发的管理

美国食品与药物管理局（Food and Drug Administration，FDA）是美国国会管辖之下的管理医疗器械的惟一联邦政府机构，它保证了美国医疗器械的安全性和有效性。FDA 早在 1906 年就开始管理食品和药物，与此不同，它对医疗器械的管理从 1938 年才开始。当时，FDA 的主要任务是确保美国市场上医疗器械的真正合法化，商标真实，没有虚假商标的产品。随着时间的推移和医疗技术的发展，FDA 所包括的医疗器械范围发生了变化。1976 年通过的医疗器械修正案（Medical Device Amendment）首次立法，提出了所有医疗器械上市销售前必须经过审查的要求。该修正案要求医疗器械新产品在上市前必须取得 FDA 的许可证，并且必须通过临床人体试验。详细内容请参考 FDA 的处理过程，其网址为 http：//www.fda.gov/。

FDA 包括五大中心：生物制品评价与研究中心、药物评价与研究中心、食品安全与营养中心、兽药中心、器械与放射卫生中心（Center for Devices and Radiological Health，CDRH）。每个 FDA 项目中心都有不同的主要管辖领域。根据 FDA 的规定，CDRH 负责确保医疗器械的安全性和有效性，并且防止医疗、工作职业和消费过程中人体可能遭受的不必要的人为辐射照射。

CDRH 下设 6 个办公室：系统与管理办公室、执法办公室、科学与技术办公室、卫生项目与工业项目办公室、监督与生物测定办公室以及器械评价办公室（Office of Device Evaluation，ODE）。其中，ODE 的主要职能如下：

1）将所有 510（k）上市前通告申请、上市前批准（Premarket Approval，PMA）、器械分类以及临床研究的器械豁免认证（Investigational Device Exemptions，IDE）通报给 CDRH 负责人。

2）对各项 510（k）、PMA 和 IDE 申请的批准、否决和撤销工作进行规划、实施和协调。

3）对申报者呈交的各种上市产品、临床技术和各种要求递交的报告进行审查、监督和医疗评估。

4）制定并解释有关器械分类以及 510（k）、PMA 和 IDE 申请的规则和指导方针。

5）参与各种国内和国际统一标准的制定。

开发销售医疗器械的厂商在产品生产之前、之中和之后都可能需要与 ODE 进行多次交涉。

原则上，如果制造商生产的产品属于医疗器械，那么就需要接受 FDA 的上市前和上市后管理（见图 2.5）。与 FDA 监管的药品等其他物品不同，根据 FDA 的规定，医疗器械的定义是：

“凡是用于人或动物的疾病诊断、治疗、缓减和预防，或者旨在对人和其他动物身体的功能和结构产生作用，而且不是通过化学反应和代谢来达到其主要功能的仪器、设备、机器、装置、植入物、外用试剂等相关物品。”

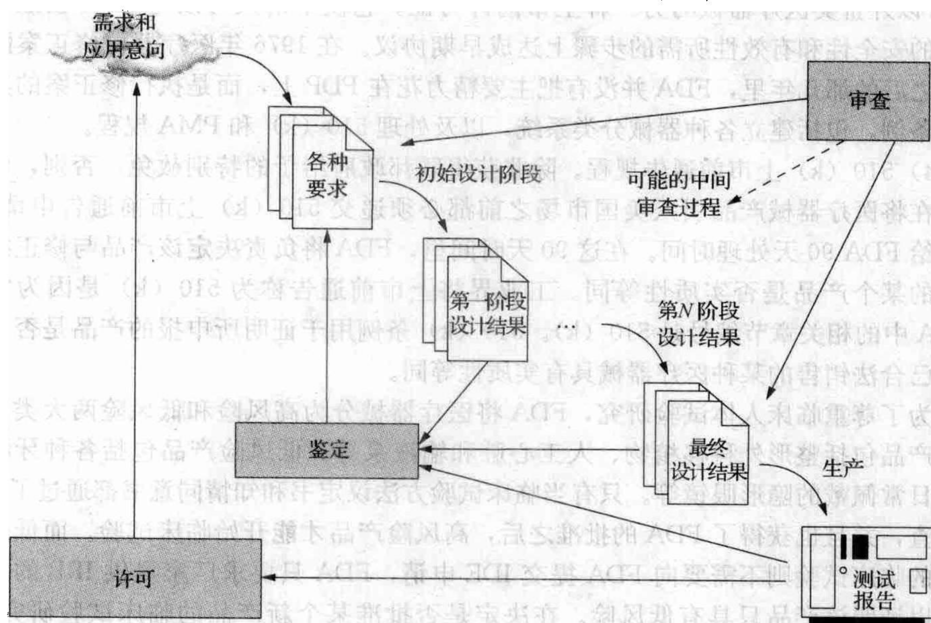


图 2.5 FDA 管理规程的目的是通过产品的审查来确保医疗器械的安全性和有效性，确保设计的质量，监管器械的质量；审查过程的结果就是对医疗器械的鉴定和许可

2.10 医疗器械的上市

在美国医疗器械的上市有如下 4 条主要途径：

1）上市前批准（PMA）。除非产品可以通过510（k）规程上市（见下述），否则具有高风险的Ⅲ类医疗器械产品的销售必须申请并通过PMA。只有FDA确定所申报的医疗器械安全且有效时，它才能获得PMA许可证。PMA申请过程可能耗资巨大，PMA所要求的数据采集费用即便不是数百万美元，也要数十万美元。而且，PMA申请的数据采集时间可能要好几年。不过，一旦获得PMA许可证，申请者就相当于获得了某种医疗器械在市场上独家销售的许可执照一样，因为别的厂家如果要销售具有同样用途的同类医疗器械也必须申请PMA批准。

2）研究性器械的豁免认证（IDE)。IDE许可规程允许厂家生产的那些只用于临床人体试验的器械得到豁免特权。由于IDE许可的产品专门用于临床试验，而不是商业化销售，因此，FDCA（即美国食品、药品和化妆品法案）授权FDA免去这些器械的某些商业化销售产品所需达到的要求。如果没有得到IDE许可，所有医疗器械的临床评估就不能作为上市的依据。IDE可以由伦理审查委员会（Institutional Review Board，IRB）单独审核，或者由IRB和FDA联合审核。

3）产品开发协议（Product Development Protocol，PDP)。PDP是除了IDE和PMA以外Ⅲ类医疗器械的另一种上市前许可证。它使申请人可以与FDA在证明新产品的安全性和有效性所需的步骤上达成早期协议。在1976年医疗器械修正案刚刚制定之后的那几年里，FDA并没有把主要精力花在PDP上，而是执行修正案的其他主要条例，包括建立各种器械分类系统，以及处理510（k）和PMA规程。

4）510（k）上市前通告规程。除非获得联邦政府给予的特别赦免，否则，任何厂商在将医疗器械产品引入美国市场之前都必须递交510（k）上市前通告申请书，并留给FDA 90天处理时间。在这90天时间里，FDA将负责决定该产品与修正案中已有的某个产品是否实质性等同。工业界将上市前通告称为510（k）是因为它在FDCA中的相关章节编号是510（k)。510（k）条例用于证明所申报的产品是否与市场上已合法销售的某种医疗器械具有实质性等同。

为了尊重临床人体试验研究，FDA将医疗器械分为高风险和低风险两大类。高风险产品包括整形外科移植物、人工心脏和输液泵等。低风险产品包括各种牙科器械、日常佩戴的隐形眼镜等。只有当临床试验方法议定书和知情同意书都通过了IRB的审查，并且也获得了FDA的批准之后，高风险产品才能开始临床试验。而低风险产品的临床试验则不需要向FDA提交IDE申请，FDA只要求厂家提供IRB的批准书，以证明该产品只具有低风险。在决定是否批准某个新产品的临床试验研究时，IRB和FDA必须确定如下几点事项：

1）受试者所承担的风险降至最低；

2）与预期的受益或者可能获得的知识相比较，受试者所承担的风险是合理的；

3）受试者的选择是公平的；

4）知情同意的材料和手续过程完整；

5）研究过程的监视和病人信息的保护等措施可以接受。

另外，FDA允许没有IDE许可证的医疗器械产品可以在以下3种特殊情况下使用，

即可行性研究、紧急使用和治疗使用。下面分别介绍这几种使用情况的伦理道德问题。

2.11　可行性研究中的伦理道德问题

可行性研究又称有限调查研究，是指只在单个机构中进行的、涉及 10 人以下受试者的新型医疗器械的人体试验。FDA 要求这种试验的申请人提交一份“有限调查研究通告书”，其中包括医疗器械的说明、调查研究的目的、试验方案、一份知情同意表以及相关医疗委员会的许可证明。在某些情况下，FDA 可能还需要其他资料，也可能需要申请人提交完整的 IDE 申请，或者暂停该项调查研究。

这种调查研究只限于：研究现有器械的新用法；研究开发初期的暂时性或永久性植入物；研究现有器械的改进。

为了充分理解这种 IDE 条例以外未批准医疗器械的临床使用可能引起的道德问题，下面我们应用本章前面所述的实践、无效实践和研究这 3 类临床人体试验的特性，来分析可行性研究应该属于哪类试验。

显然，可行性研究的目标是获取知识，因此，它是一种研究而不是实践。厂家希望通过特定病人群体的使用来考察器械的性能，从而获得其有效性和安全性的信息。这些信息对于决定该器械是否需要进行深入的动物或人体研究，是否还需要改进等，具有重要意义。未批准医疗器械的可行性研究与 IDE 许可的应用，这两者之间的主要区别是 FDA 对前者的审查力度远小于后者。这就意味着 IRB 对保证器械可行性研究应用的合理性承担了主要责任。

这里首先必须弄清楚的道德问题是，使研究合理化的首要因素是什么。毕竟，无论某个医疗器械进行了多少基础研究和动物实验，在还没有实际用于人体时，它对人体的风险和益处都不可能完全确定。人体研究的好处主要在于能够获取信息，这些信息对于开发既有效又安全的医疗新方法至关重要。因此，人体试验研究合理化的一个伦理条件就是它必须具有科学上的合理性。

虽然科学合理性是人体试验的必要道德条件，但它并不充分。在这种试验中受试者有可能成为纯粹的研究资源，也就是只有研究价值。人不能只有被使用的价值，他们的价值就在于他们是人，应该受到尊重，也就是要尊重人的自主权。要做到这一点，就必须保证，如果事先没有给出自愿的知情同意，任何人都不能参加临床试验。而且，尊重人就意味着医生不能让病人承担不必要的风险，也就是在试验过程中要尽可能减少病人的风险。

FDA 在审查 IDE 条例的未批准医疗器械的使用申请时，主要关注研究的两个伦理条件：一是试验的科学合理性，二是尊重受试者的权利。医学伦理学家指出，如果削弱 FDA 的审查力度，就会增加不满足这两个条件的事件的发生概率。因为医疗器械生产厂家毕竟是商业企业，他们追求的是利润，他们会尽量减少时间和费用，尽可能快地将他们的产品推入市场。这些自身利益的动机经常会与伦理要求发生冲突，从而使得厂家不能满足这些伦理要求。而且，利润并不是驱使厂家违反人体试

验伦理要求的惟一动机，厂家可能会真诚地认为他们的产品会给许多人带来巨大的利益，从而在这种思想的激励下选择某些捷径，降低人体试验研究质量的要求。无论人体试验是为了自身利益还是他人利益，道德问题都是相同的。必须避免将受试者当做物体而不是人来承担风险。

那么，开展可行性研究有怎样的条件要求呢？这些要求是否足以使其有别于“正常”条件下的研究而可以减少 FDA 的审查呢？正如以上所述，厂家希望通过可行性研究来考察现有器械的新用途、测试处于开发初期的临时性或永久性植入物的性能、或者研究现有器械的改进。前面也已经提到，可行性研究只能在一个机构中进行，且受试者人数不能超过 10 人。那么，在这些条件下，可行性研究的科学性和对人权的尊重是否会低于“正常”人体研究的水平呢？

可行性研究只在很小的受试人群中进行，与一般人体试验相比，其道德失误产生的伤害只影响少数人。但是，即使其伤害只限于 10 人或更少的受试者，这种伤害仍然是一种道德错误。尽管伤害 10 人以下的问题不像伤害 10 人以上那么糟，但是，无论如何这是不道德的行为。

那么，可行性研究中发生的道德失误是否可能比执行 IDE 条例的要多呢？虽然前面的论述中没有明确回答这个问题，但是，这是 FDA 应该优先给予重视的问题。是临时性或永久性植入物，还是已批准器械的新用途开发或改进；对于不同类型的医疗器械，这个问题的答案可能截然不同。不过，无论可行性研究的预期应用是什么，除非有足够的理由能够确认受试者不会承担比“正常”试验更大的风险，否则，在道德上 FDA 不应该允许这种未批准器械的非 IDE 使用。

2.12 紧急使用的伦理道德问题

医疗器械的紧急使用可以避开 IDE 的严格审查，其中的机制是什么呢？如果某个未批准器械是挽救濒死病人的惟一选择，那么，即使该器械的使用还没有获得 IDE 许可证，或者 IDE 许可已通过，但是想使用该器械的医生却不是 IDE 许可的研究人员，在这些情况下，FDA 允许该器械的紧急使用。

允许未批准器械的紧急使用是为了在没有其他可用方法的情况下挽救濒死病人的生命。这种使用属于实践而不是研究。它主要是为了病人的利益，而不是为了获取新知识。由于器械在还没有完成临床试验之前就使用了，因此，这种使用属于无效实践。这是什么意思呢？

首先，虽然器械的使用目的是为了挽救病人的生命，但是，与那些属于有效实践的临床应用相比，这种器械使用可能带来的潜在益处和风险要不确定得多。有效实践是指已经确立了有效性和安全性的器械的使用，它们经过全面的调查研究，包括临床前研究、动物实验和人体试验；因此，临床医生完全有依据来判断这种治疗方法对于病人的益处和风险。

而无效实践所缺少的正是这样的依据。那么，这是否意味着未批准器械的紧急

使用是不道德的呢？如果没有任何依据可以判定器械的使用将带来什么风险和益处，那么，就可以认为其使用是不道德的。不过，FDA 要求施行未批准器械紧急使用的医生必须具有充足的理由相信该器械的使用会有好处，这就必须有充分的临床前研究和动物实验的结果作为依据来预测器械对病人的益处。

由此可见，虽然这种器械的使用所带来的益处和风险相当不确定，但并不是完全不确定。尽管医疗新技术合法化的惟一途径需要经过临床人体试验，但并不是所有无效实践的无效程度都是一样的。有些医疗技术基本上没有进行深入的研究，其益处和风险几乎完全不能确定；而有些医疗技术至少已经有临床前研究和动物实验数据的支持，虽然这些数据还不足以支持器械的常规临床应用，但是它们可以作为紧急情况下铤而走险的措施的依据。不过，这种铤而走险的措施绝不是不计后果的行为，因此，在道德上，FDA 首先要求未批准器械的紧急使用必须具有临床前研究和动物实验结果的有力支持。

FDA 对于未批准器械紧急使用的第二个要求是医生不能事先断定紧急情况的存在。FDA 希望医生对于潜在的紧急情况进行合理的预测，并根据 IDE 规程安排有关器械的使用，而不是随便实施紧急使用。这样，医生就不能为了逃避 IRB 的审查，为了避免申请未批准器械的使用权，而"制造"紧急事件。从康德主义保护人的尊严的观点来看，这个要求特别重要。制造紧急事件来逃避 FDA 的管理，这是欺骗病人，使病人成为纯粹的试验资源，将病人的价值降低为只为临床医生的目标服务。因此，FDA 坚持紧急情况是指那些无法合理预测的情况，这是非常正确的。

这里，病人的知情同意同样非常重要。面临死亡的病人尤其容易被利用，应该给予更多的保护。措施之一就是要确保病人或者病人的合法监护人了解即将使用的治疗方法存在高度的不确定性，使他们清楚地认识到，与那些有效实践相比，临床医生估计这种新疗法的风险和益处的依据是相当不充分的，这种即将使用的器械还没有经过严格完整的人体试验，医生对器械潜能的估计只是基于临床前研究和动物实验的结果。还有一点也很重要，就是要让病人知道所有的可能性，不仅只是生或死，还有质量严重受损的生存。虽然紧急情况下可以合法地采取铤而走险的措施，但是是否采取这种措施必须由知情同意的病人自愿决定，当然这是特别虚弱的病人。

此外，参与未批准器械的紧急使用的医生必须认识到这种行为虽然是无效实践，但属于实践性的，不是研究性的。因此，其行为的首要目标就是病人的幸福。这种情况下病人与临床医生之间的信任关系与正常临床治疗情况下相同。如果将这种情况看做研究性的，以其对科学和社会的益处为目标，那么这是对病人与医生之间信任关系的亵渎。

案例 5：医疗专家系统

很多学科都开发了各种专家系统，临床决策系统就是一个例子，这种系统用于模拟医生的决策技能。不过，系统的性能取决于一种公认的知识体系的建立，

该知识体系体现了医生在特定的具体数据情况下所采取的决策途径。临床决策系统作为一种咨询系统，可以给医生提供建议，但最终作决定的还是医生。

假如有一个用于精神病药物治疗监控的系统，该系统在临床医生的协助下由生物医学工程师设计完成。如图 2.6 所示，给系统输入病人的具体诊断结果之后，就可以得到治疗某种精神病的合适的推荐药物。医生选择某种推荐药物，并选定用药剂量，以确定特定病人的药效。治疗期间通过验血确定药物的浓度，并通过其他心理检查来判断药物是否达到了理想的作用效果。

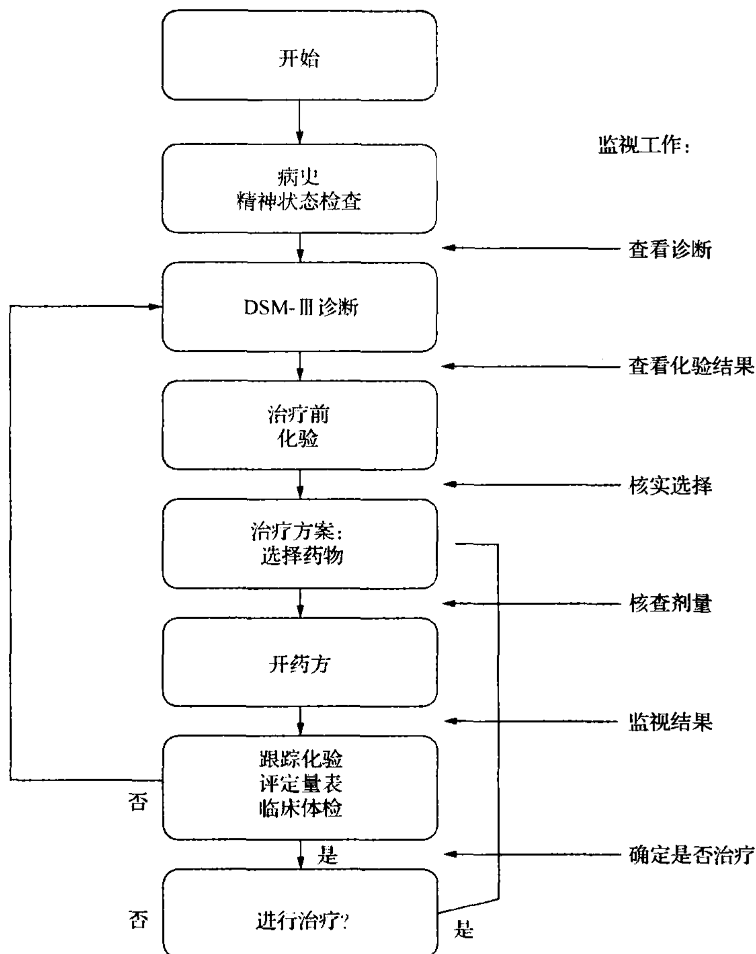


图 2.6 临床医生实行药物治疗过程的流程

然后，将检查获得的数据输入决策系统，系统会将这些数据与预期的标准结果进行比较，如果数据超出预期的极限，系统就会给出警报信息，提醒医生必须进一步采取措施。

这种临床决策系统的应用存在如下问题：

1）谁将对错误治疗负责，是医生、编程人员，还是系统管理人员？

2）是什么造成了错误治疗？

3）系统设计人员的职责是什么？换言之，成功的设计由什么组成？

4）临床上如何评价使用这种系统的医生的工作，又如何评价这种系统本身？

2.13 治疗使用的伦理道德问题

当已批准药物对病人毫无疗效时，FDA 的管理条例允许医生使用正在研究的新药。这种未批准新药的使用并不只限于生命受到威胁时的紧急情况，还可以用于治疗严重疾病。但是与药品不同，FDA 不允许未批准医疗器械的治疗使用，不过，厂家如果在 IDE 申请书中提供特殊的治疗使用方案议定书，就可能获得这种治疗使用的许可。

未批准医疗器械的治疗使用的准则与研究性新药的治疗使用相似，它包括：该器械旨在治疗危重疾病或者威胁生命的疾病；没有其他功能相当或者令人满意的产品可供使用；该器械正处于 IDE 审查中，或者已获得 IDE 赦免认证，或者已完成所有临床试验正在等待批准；厂家正在积极进行该研究性器械的上市销售申请。

治疗使用的方案议定书要作为 IDE 申请中的一部分上交，并要求说明器械的用途、基本原理、现有的替代品、本器械的优势、适应症器械使用情况的监测方法和降低风险的方法以及有关器械安全性和有效性的技术信息。

那么，FDA 应该允许未批准医疗器械的治疗性使用吗？这会带来什么道德问题呢？首先，这种治疗性使用的前提是现有的已批准治疗方法无法改善病人的状况，其使用是为了病人的治疗，因此它属于实践，而不是研究。其次，所用器械还未获得批准，因此，这种使用属于无效实践。这样，与紧急使用情况一样，治疗性使用也必须满足 FDA 的要求，这就是必须具有充分的临床前研究和动物试验的结果可以证明对病人有益。与紧急使用相同，虽然这些研究结果并不能排除该产品在利益和风险上的高度不确定性，但它们其实排除了完全不确定性。同样，虽然在绝望的状况下可以采取铤而走险的措施，但这绝不能是不计后果的行为。另外，与紧急使用不同，治疗使用的对象是健康受到严重损害的病人，而不是濒临死亡的病人。因此，这里必须考虑的问题是，健康损害达到怎样的严重程度才可以求助于这种风险和益处还没有充分确定的治疗方法。

对于紧急使用，FDA 要求医生不能利用这种使用方式来人为“制造”IDE 规程的例外，从而逃避有关审查规程。同理，接受治疗使用的病人也是特别容易受到伤害的病人，虽然他们还不至于面临死亡，但他们面临的是严重的疾病，他们比境况较好的病人更容易被利用。因此，必须让病人了解治疗效果的不确定性，也就是这种治疗可能对病人没什么益处。这一点非常重要。

2.14 FDA认证过程中生物医学工程师的职责

1991年11月28日，美国1990年医疗器械安全法（Safe Medical Devices Act of 1990）（即公法101-629）开始生效。该法规要求广大医疗保健机构，包括医院、流动性外科服务站、疗养院和门诊医疗机构，通报医疗器械引起的病人死亡、严重受伤和严重患病的情况。其中，与医疗器械相关的死亡事件必须同时向FDA和器械生产厂家通报；与医疗器械相关的严重患病和受伤事件则必须向厂家通报，万一不知道厂家的联系方式，就向FDA通报。此外，上述医疗保健机构每半年必须向FDA提交一份这半年期间已上交报告的总结。在这项法规出台之前，这种报告完全是自愿提交的。新法规旨在使FDA能够迅速了解与医疗器械相关的问题，它其实是1984年颁布的医疗器械报告条例（Medical Device Reporting，MDR）的补充。MDR只要求厂家和进口商向FDA提交与医疗器械相关的死亡和严重受伤事件的报告，而新法规将该要求的范围从厂家和进口商扩大到了医疗器械的用户，也就是新法规赋予FDA管理医疗器械用户的权力。

FDA的这种管理模式具有深刻的道德意义，通过增进FDA对于医疗器械相关问题的了解，可以增强政府机构保护病人福利的能力。对于FDA管理政策的主要争议是其功利主义本质。该法规的怀疑者不相信它能够给FDA提供较多的有用信息，他们担心这一新法规带来的很多信息可能只是复制以前MDR法规下的有关信息。如果真是如此，该法规没有给病人带来什么利益。此外，根据那些怀疑者的言论，这些规定很可能增加病人对医院和厂家提出的诉讼，并且，医疗器械用户还需要设立正规的系统来报告设备相关的问题，需要安排人员运行这种系统；显然，这会增加医疗保健的成本，从而加剧就医难的问题，很多人认为就医难已经成为一种危机。总之，对于FDA政策的争议焦点是担心病人的利益会被边缘化，而就医成本的负担则会加重。

生物医学工程师必须了解FDA有关各种医疗器械和系统使用的规定和审批程序，这些管理政策实际上是控制医疗器械不正当使用的社会机制。

习　题

1. 请解释伦理和道德这两个术语之间的区别，并分别举例说明医学中的伦理和道德问题。

2. 请列举3个医疗道德评判的例子。

3. 功利主义的支持者们相信怎样的学说？

4. 康德主义所期望的病人的权利和意愿是什么？

5. 请论述如何应用临床工程师的道德准则来指导具体工作。

6. 请论述脑死亡的含义，它与新皮质死亡有什么区别？

7. 主动安乐死与被动安乐死、自愿安乐死与非自愿安乐死的区别是什么？你认为哪种安乐死应该被准许？请说明理由，并说明需要满足的条件。

8. 如果某位住在重病监护室的病人的家属递交了病人的遗嘱，那么应该立即执行遗嘱还是医生和家属应该讨论一下？谁应该作出决定？为什么？

9. 什么是人体试验？什么情况下允许进行人体试验？医院应该采取哪些保护措施？

10. 假设生物医学工程师设计了一种新型睡眠呼吸暂停监视仪，请讨论将其用于临床之前必须经过哪些处理程序。

11. 请论述实践、研究和无效实践之间的区别，并列举它们在医学领域中的例子。

12. 符合伦理道德的研究需要满足哪两个主要条件？

13. 知情同意是允许人体医学试验的主要条件之一。知情同意满足的是什么道德原则？怎样才能保证受试者真正的自愿性？受试者应该知道哪些信息？

14. 可行性研究和紧急使用之间有什么区别？

15. 在医疗实践中，医务人员需要使用医疗器械进行病人的诊断和治疗；因此，他们不仅需要具备人体生理学的知识和经验，而且还必须能够使用他们所支配的医疗器械。医疗器械事故经常会导致病人的起诉。显然，这里存在的问题是谁应该对医疗器械事故负责。

假如有这么一个案例，一位妇女手术时需要使用一块接地板，给手术中所用电气设备产生的电流提供回路。接地垫板一般为 8in[⊖] × 11in。结果，这位妇女身上紧贴接地板的皮肤组织被严重烧伤。

(a) 请讨论哪些个人和组织可能对这起事故负有责任；

(b) 除了寻找合理的责任承担者以外，这里是否存在特殊的伦理道德问题？

推荐阅读资料

Abrams, N. and Buckner, M.D. (Eds.) (1983). *Medical Ethics*. MIT Press, Cambridge, MA.

Bronzino, J.D., Smith, V.H. and Wade, M.L. (1990). *Medical Technology and Society*. MIT Press, Cambridge, MA.

Bronzino, J.D. (1992). *Management of Medical Technology*. Boston, Butterworth, 1992.

Chapman, A.R. (1997). *Health Care and Information Ethics: Protecting Fundamental Human Rights*. Sheed and Ward, Kansas City, KS.

Dubler, N. and Nimmons, D. (1992). *Ethics on Call*. Harmony, New York.

Jonsen, A.R. (1990). *The New Medicine and the Old Ethics*. Harvard Univ. Press, Cambridge, MA.

Moskop, J.C. and Kopelman, L. (Eds.) (1985). *Ethics and Critical Care Medicine*. Reidel, Boston.

Pence, G.E. (1990). *Classic Cases in Medical Ethics*. McGraw-Hill, New York.

Rachels, J. (1986). *Ethics at the End of Life: Euthanasia and Morality*. Oxford Univ. Press, Oxford.

Reiss, J. (2001). *Bringing Your Medical Device to Market*. FDLI Publishers, Washington DC.

Seebauer, E.G. and Barry, R.L. (2001). *Fundamentals of Ethics for Scientists and Engineers*. Oxford Press, New York.

⊖ 1in=25.4mm。

第 3 章　解剖学与生理学

Susan Blanchard 博士

本章目录

本章主要内容：

- 解剖学和生理学的定义，以及它们对于生物医学工程的重要性；
- 重要解剖学术语的定义；
- 细胞理论；
- 细胞中的几种主要有机物等成分；
- 细胞膜保持细胞体积及其内容物浓度的机理；
- 平衡状态下细胞模型内部渗透浓度和离子浓度的计算方法；
- 哺乳动物细胞内主要细胞器的功能；
- 遗传信息复制、转录和翻译之间的相似点、不同点以及它们的功能；
- 循环系统、呼吸系统、神经系统、骨骼系统和肌肉系统，这 5 大器官系统的主要组成部分及其功能；
- 自身平衡及其维持的反馈机制。

3.1　绪论

生物医学工程学是建立在工程学和生命科学基础上的交叉学科，从事生物医学工程的人员必须掌握这两个领域的知识，并能够融会贯通；必须充分了解人体的基本组成及其功能，以便与生理学家和生命科学家进行思想交流和信息互通。解剖学和生理学是生命科学最基础的两门学科，其中，解剖学介绍人体内部和外部的结构及其相互之间的物理关系，生理学则介绍各种人体结构的功能。

图 3.1a 所示是男性人体的解剖学姿势：人体直立，面向前方，双臂自然下垂在身体两侧，手掌向前。该图是前视图（即腹侧图）。而图 3.1c 所示是处于解剖学姿势的男性人体的后视图（即背侧图）。图 3.1b 则是女性人体的侧视图。临床上用方向术语来表示人体各个不同部位之间的相对位置关系。例如，“近端”和“远端”是表示某部位与躯干或相连肢体之间的距离的远近（见图 3.1a）；“上”表示解剖方位中靠近头部的部位，“下”则表示靠近足部的部位（见图 3.1b）；“内侧”表示靠近身体中轴的部位，“外侧”则表示远离中轴的部位（见图 3.1c）。处于头部方向的身体部分称为头侧，处于足部方向的身体部分则称为尾侧（图 3.2）。

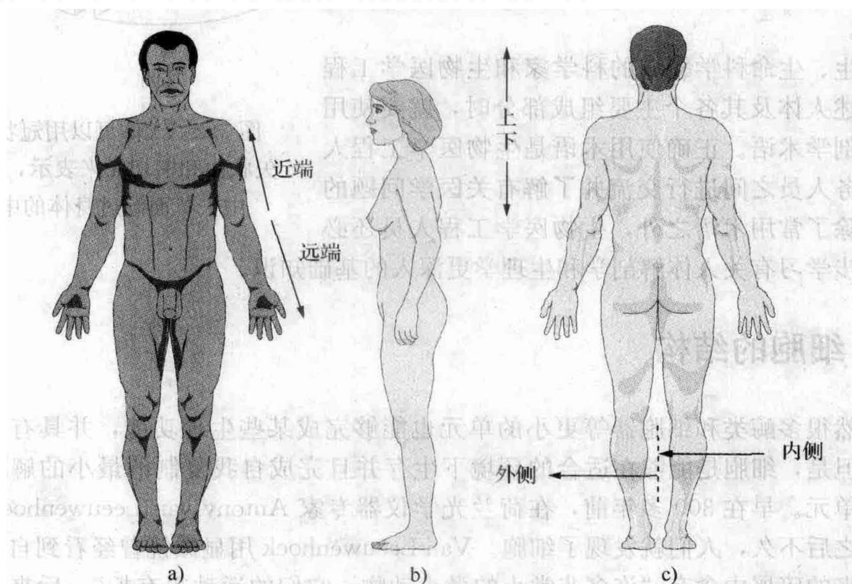


图 3.1　解剖学姿势

a）男性人体的前视图　b）女性人体的侧视图

c）男性人体的后视图（图中显示了近端和远端、上和下，以及内侧和外侧等各种相对的方向）

解剖位置也可以用平面来表示。沿着中线将人体分为左右对称两半的平面称为中矢状面（见图 3.2），平行于中矢状面的其他非对称分割身体的平面称为矢状面；垂直于矢状面且前后平分身体的平面被称为冠状面；既垂直于矢状面又垂直于冠状

面的横切身体的平面称为横切面。

人体分成主干和四肢两大部分，主干包括头、颈、胸、腹和骨盆，四肢则包括上肢和下肢。其中，上肢包括肩、上臂、前臂、腕和手，下肢包括髋、大腿、小腿、踝和足。腹部还可以进一步分为 9 个分区或者 4 个分区。

体腔用于容纳身体内部的各种器官，主要有背侧体腔和腹侧体腔，还有一些较小的腔室，包括鼻腔、口腔、眼窝、中耳鼓室以及活动关节的关节腔等。其中，背侧体腔包括容纳脑的颅腔和容纳脊髓的脊髓腔。腹侧体腔则包括由横膈膜分开的胸腔和大腹腔。胸腔由肺和纵膈组成，纵膈包括心脏及其血管、气管、食道和除肺以外的该部位的所有其他器官。大腹腔被人为地分为腹腔和盆腔两个部分。腹腔是身体最大的腔室，容纳胃、小肠、大肠、肝、脾、胰、肾脏和胆囊；盆腔则容纳膀胱、直肠和内生殖器。

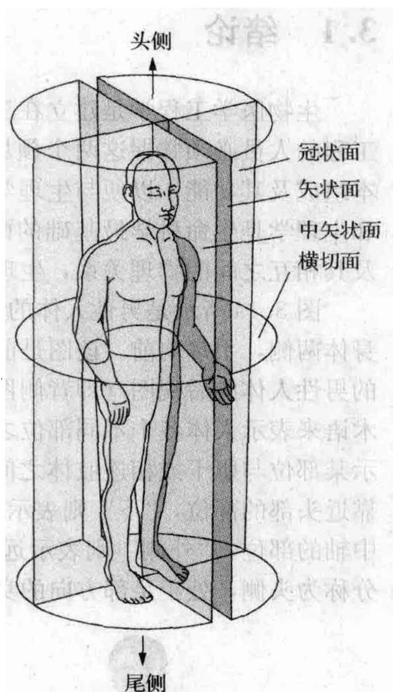


图 3.2 人体可以用冠状面、矢状面和横切面来表示，其中，中矢状面经过身体的中轴

医生、生命科学领域的科学家和生物医学工程人员表述人体及其各个主要组成部分时，就要使用这些解剖学术语。正确使用术语是生物医学工程人员与医务人员之间进行交流并了解有关医学问题的基础。除了常用术语之外，生物医学工程人员还必须进一步学习有关人体解剖学和生理学更深入的基础知识。

3.2 细胞的结构

虽然很多酶类和细胞器等更小的单元也能够完成某些生理功能，并具有特定的结构，但是，细胞是能够在适合的环境下生存并且完成自我复制的最小的解剖单元和生理单元。早在 300 多年前，在荷兰光学仪器专家 Antony van Leeuwenhoek 发明显微镜之后不久，人们就发现了细胞。Van Leeuwenhoek 用显微镜曾经看到自己牙垢中刮下来的碎屑中含有“许多非常小的微小动物，它们的运动很有趣”。后来，到了 17 世纪后期，英国皇家仪器学会管理员 Robert Hooke 用最早发明的一台显微镜观察到了软木塞中的植物细胞的细胞壁，并描述了细胞的形态。根据这些观察以及其他研究结果，Theodor Schwann 和 Matthias Jakob Schleiden 提出了细胞理论，并在 19 世纪中期由 Rudolf Virchow 完善了这一理论。细胞理论指出：任何生物体都由一个或多个细胞组成；细胞是生命的最小单位；任何细胞都来自先前存在的细胞。因此，细胞是生命的基本结构单元。

细胞主要由有机化合物和水组成，人体体重至少有 60%是水。碳水化合物、脂类、蛋白质和核酸等细胞合成的有机化合物分子是维持生命的基础，这些分子具有存储能量、传递遗传信息、提供结构材料以及完成代谢等功能。按照所占体重百分比降序排列，人体中最常见的化学元素是氧、碳、氢、氮、钙、磷、钾、钠、氯、镁、硫，铁和碘。其中，碳、氢、氧和氮在所有组成人体的元素中所占的比例超过 99%，这些元素大部分存在于有机化合物中，但有一些以磷酸基团等其他离子形式存在。

细胞的各种有机化合物具有不同的特性和功能。碳水化合物不仅在细胞中用作结构材料，而且还用于运输和存储能量，它有 3 种存在形式：单糖（如葡萄糖等）、双糖（如乳糖、蔗糖、麦芽糖等）和多醣（如糖原等）。脂类是油性化合物，互相之间可以溶解，但不溶于水。它们是构成细胞的材料，也是体内能量的主要存储形式。蛋白质是生物分子中形式最多样化的物质。酶就是一种特殊的蛋白质，它可以加快新陈代谢的反应速率，使细胞生产出生命所需的各种有机化合物。其他各种蛋白质，有的是细胞的结构元素、有的是跨膜运输通道、有的是改变细胞活动的信号、还有的是对抗携病细菌的化学武器。各种各样的蛋白质都由 20 种基本氨基酸构成。

核苷酸和核酸是这里要介绍的最后一类重要生物分子。核苷酸是一种较小的有机化合物，它由一个五碳糖（即核糖或脱氧核糖）、一个磷酸基团以及一个含单个或两个碳环结构的含氮碱基组成。三磷酸腺苷（即 ATP）是一种核苷酸，它是细胞的能量货币，在新陈代谢中起着核心作用。还有一些核苷酸是协助酶发挥功能的辅酶的亚基。核酸有两种，即脱氧核糖核酸（DNA）和核糖核酸（RNA）。如图 3.3 所示，DNA 是一种独特的螺旋状分子，含有两条方向相反的配对核苷酸链，其中，每个核苷酸要么是单环结构的嘧啶碱基——胸腺嘧啶（T）或胞嘧啶（C），要么是双环结构的嘌呤碱基——腺嘌呤（A）或鸟嘌呤（G）。在 DNA 的双螺旋中，胸腺嘧啶总是与腺嘌呤配对，即 T-A；而胞嘧啶总是与鸟嘌呤配对，即 C-G。RNA 与 DNA 相似，但它是单螺旋结构，且组成 RNA 的是核糖而不是脱氧核糖，另外，其嘧啶碱基不用胸腺嘧啶，而用尿嘧啶（U）。

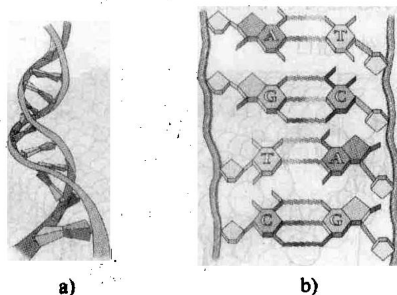


图 3.3　a）DNA 由两条方向相反的配对核苷酸链组成，形成螺旋结构　b）通过碱基之间氢键的结合，胸腺嘧啶与腺嘌呤配对，即 T-A；而胞嘧啶与鸟嘌呤配对，即 C-G

任何细胞都由细胞膜包围着，使其与外界环境分开，但内外环境之间并不是绝对隔离的。动物细胞，如人体细胞，都是真核细胞。如图3.4所示，真核细胞除了细胞膜之外，还包含有质膜包裹的各种细胞器和一个细胞核。而原核细胞（如细菌等）除了细胞膜之外就没有其他质膜包裹的结构了。此外，任何细胞都含有DNA（携带细胞遗传信息）和细胞质（质膜内除DNA外的所有半流体物质）。

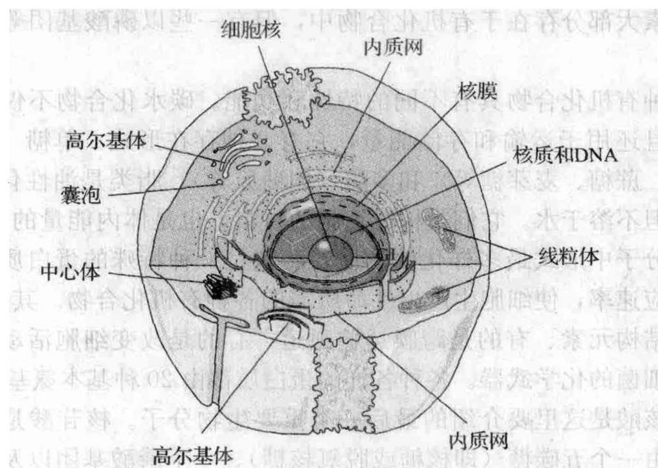


图3.4 动物细胞由细胞膜包围，其中含有一个由质膜包裹的细胞核，核内有DNA分子；细胞核外的细胞质中含有具有特殊功能的细胞器

3.2.1 细胞膜

细胞膜为细胞提供了许多功能。它具有机械强度，能够支撑细胞的结构，协助细胞运动，并通过调节进出细胞的化学物质来控制细胞的体积及其活性。如图3.5所示，细胞膜由磷脂双分子层以及其中结合的蛋白质和胆固醇等组成。哺乳动物细胞的细胞膜蛋白质含有激素的结合位点、区分细胞的识别位点、相邻细胞相互粘附的位点以及跨膜物质转运的通道等。磷脂分子都排列成“亲水”头部向外，“疏水”

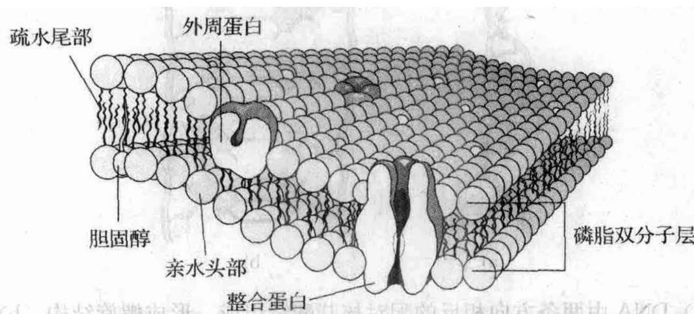


图3.5 任何细胞都由细胞膜包围。细胞膜由结合了蛋白质和胆固醇的磷脂双分子层组成

尾部向内。这种磷脂双分子层上结合了蛋白质通道，可以控制各种物质的跨膜运输，从而维持细胞内环境的稳定。磷脂双分子层上结合的胆固醇分子就像稳定剂，防止脂质分子沿膜平面的过度侧向运动。

氧、二氧化碳和水等分子可以轻易地穿过细胞膜，而其他大分子和离子等就必须经过蛋白质通道运输。渗透是物质穿过细胞膜等选择性通透膜的一种方式，也就是分子从相对的高浓度区域扩散到低浓度区域的过程。当物质跨膜流动的净通量为零，即这种物质在细胞内外的浓度相等时，才能够达到扩散平衡。渗透是一种被动运输过程。主动运输可以从低浓度区域逆浓度梯度将离子和分子跨膜运输到高浓度区域，它需要消耗能量，通常由 ATP 提供。这种运输机制维持了细胞内外的离子浓度差。一般，哺乳动物细胞胞内钠离子（Na^+）浓度是 12mmol/L，即每 1000 升溶液中含有 12mol 的 Na^+，胞外 Na^+ 的浓度是 120mmol/L；细胞内外钾离子（K^+）浓度分别为 125mmol/L 和 5mmol/L。除了带正电荷的离子（即阳离子）以外，细胞也含有带负电荷的离子（即阴离子）。例如，哺乳动物细胞内外的氯离子（Cl^-）浓度分别为 5mmol/L 和 125mmol/L。胞内其他阴离子，如蛋白质、带电荷的氨基酸、硫酸根离子和磷酸根离子等，总浓度约为 108mmol/L。这些跨膜离子浓度梯度可用于生产 ATP，驱动各种运输过程，并且产生电兴奋信号。

例 3.1　体积为 2nl 的细胞内含有多少个钠离子和钾离子？

解：假设细胞内 Na^+ 和 K^+ 的浓度分别为 12mmol/L 和 125mmol/L，则根据细胞的体积和阿伏伽德罗常数，就可以计算出这两种离子的数目，即

$$Na^+：\frac{12\text{mol}}{1000\text{L}}\times 6.023\times 10^{23}\frac{\text{分子}}{\text{mol}}\times 2\times 10^{-9}\text{L}=1.45\times 10^{13}\text{分子}$$

$$K^+：\frac{125\text{mol}}{1000\text{L}}\times 6.023\times 10^{23}\frac{\text{分子}}{\text{mol}}\times 2\times 10^{-9}\text{L}=1.51\times 10^{14}\text{分子}$$ ■

细胞膜通过控制细胞内的渗透浓度（osmolarity）来调节细胞的体积。渗透浓度是指溶液中游离微粒的浓度。渗透浓度为 1Osm 的溶液表示 1L 溶液中含有 1mol 微粒，而 1mOsm 渗透浓度则表示 1000L 溶液中含有 1mol 微粒。高渗透浓度的溶液中水等其他溶剂的浓度较低。在生物学中，0.1Osm 葡萄糖和 0.1Osm 尿素溶液中所含的水浓度相同。必须注意，0.1mol/L 氯化钠（NaCl）溶液的渗透浓度是 0.2Osm，因为 NaCl 在溶液中会分解成 Na^+ 和 Cl^- 两种离子，因此其溶液中的微粒要比葡萄糖等不能分解的物质多一倍。如果两种溶液的渗透浓度相等，那么这两种溶液等渗；如果一种溶液的渗透浓度低于另一种溶液，那么这种溶液被称为低渗溶液；反之，则称为高渗溶液。注意，溶液的等渗、低渗和高渗只取决于那些不能自由跨膜流动的微粒，对于可以自由通过细胞膜的微粒，它们最终会达到平衡，即细胞内外的浓度相等。

设有一个简化细胞模型，由细胞膜和细胞质组成，胞内只含有不能穿过细胞膜的蛋白质和能够跨膜流动的水分子。平衡状态下，细胞内外的渗透浓度必须相等。因为，如果细胞内外渗透浓度不同，水分子就会从浓度高的一侧流向浓度低的一侧，

直到平衡为止。例如，图 3.6 所示的细胞模型含有 0.2mol/L 蛋白质，将其放入含有 0.1mol/L 蔗糖的低渗溶液中。设细胞膜对蛋白质和蔗糖不通透，但水分子可以自由进出。并且，假设细胞体积为 1nL，相对于胞外溶液的体积，它非常小，也就是细胞体积的变化对于胞外溶液体积的影响可以忽略不计。那么，达到平衡状态时，细胞的体积将是多少？

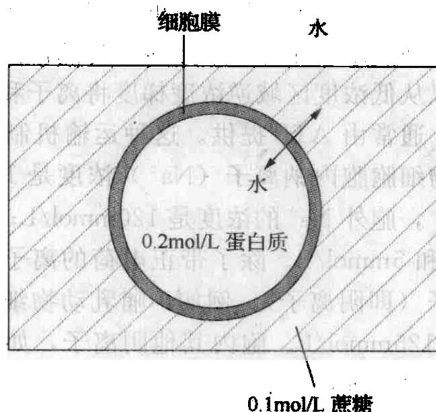


图 3.6　由细胞质和细胞膜组成的简化细胞模型

注：其细胞质中含有 0.2mol/L 蛋白质，细胞置于浓度为 0.1mol/L 的蔗糖溶液中，细胞膜对蛋白质和蔗糖都不通透，只有水可以自由出入，胞外溶液的体积比细胞体积 1nL 大得多，图中只显示了一部分。

平衡时，细胞内外的渗透浓度必须相等。由于蛋白质不能分解，细胞内的起始渗透浓度就是 0.2Osm，细胞外蔗糖溶液的渗透浓度为 0.1Osm。0.2Osm 的溶液每升中含有 0.2mol 微粒，而同等体积的 0.1Osm 溶液所含微粒只有 0.2Osm 溶液的一半，因此，细胞内的渗透浓度降低 1 倍才能达到平衡。由于细胞膜不能通透蛋白质分子，其含量不会变化，因此，只有使细胞体积扩大 1 倍才能使胞内渗透浓度降低 1 倍。于是，水分子通过细胞膜进入胞内，直至细胞体积增加到 2nL 为止。从而使得胞内渗透浓度降为 0.1Osm，与胞外渗透浓度相等。在此过程中，水分子顺着浓度梯度，从胞外含水浓度高的 0.1mol/L 蔗糖溶液扩散到胞内含水浓度低的 0.2mol/L 蛋白质溶液。

例 3.2　如果将图 3.6 所示的细胞模型放到纯水中，那么细胞将如何变化？

解：水分子能够跨膜流动，因此，它会顺着浓度梯度从水浓度高的胞外空间流向水浓度低的胞内空间。并且，只要细胞内存在蛋白质，胞外的水分子浓度总是高于胞内的水分子浓度，因此，水分子一直会流入细胞，直到细胞膜被撑破为止。■

例 3.3　假设图 3.6 所示细胞模型的初始体积为 2nL，并且含有 0.2mol/L 蛋白质。细胞置于 0.2mol/L NaCl 的大体积溶液中，假设 Na^+ 和 Cl^- 都不能穿过细胞膜进入细胞。那么，0.2mol/L NaCl 溶液相对于细胞内部的渗透浓度是低渗、等渗、还是高渗？并说明细胞在此新环境中达到平衡时发生了什么变化，细胞最终的渗透浓度和体积分别是多少？

解：细胞内的渗透浓度为 0.2Osm，而胞外 0.2mol/L NaCl 溶液的渗透浓度为 0.4Osm，即 0.2Osm Na^+ +0.2Osm Cl^-。因此，相对于细胞内的渗透浓度，NaCl 溶液是高渗的，即胞外渗透浓度大于胞内渗透浓度。由于蛋白质、Na^+ 和 Cl^- 都不能跨膜传输，因此，水分子会从胞内流出，直到胞内渗透浓度达到 0.4Osm 为止，此时胞内体积将从 2nL 减小到 1nL，即

$$C_1V_1=C_2V_2$$

$$V_2=\frac{0.2\text{Osm}}{0.4\text{Osm}}\times 2\text{nL}$$

$$V_2=1\text{nL}$$

■

实际细胞要比上述简化模型复杂得多，除了要达到渗透平衡以外，还必须根据细胞内的离子情况，达到电荷平衡。根据电中性原则，细胞等生物体的房室中，阳离子的总浓度必须等于阴离子的总浓度。例如，如图 3.7 所示的细胞模型，其细胞内外阴阳离子的浓度与一般哺乳动物细胞相近。如果细胞膜只允许 K^+ 和 Cl^- 自由通过，但对 Na^+ 和胞内其他阴离子不通透，那么，这个细胞处于平衡状态吗？我们来分析一下。细胞内的总渗透浓度为 250mOsm，即 12mmol/L Na^+、125mmol/L K^+、5mmol/L Cl^-、108mmol/L 其他阴离子的总和；而细胞外总渗透浓度也是 250mOsm，即 120mmol/L Na^+、5mmol/L K^+、125mmol/L Cl^- 的总和，因此细胞处于渗透平衡，也就是进出细胞膜的水分子净流量为零。如果胞内除 Cl^- 以外的每个阴离子所带平均电荷为 −1.2，那么细胞内也基本处于电荷平衡状态，即 Na^+ 和 K^+ 的正电荷之和为 12+125；Cl^- 和其他阴离子的负电荷之和为 5+1.2×108，两者基本相等。不过，由于实际细胞允许通透少量的 Na^+，因此，细胞要维持这种平衡就必须消耗能量，逆着 Na^+ 的扩散方向和电位梯度，通过特殊的蛋白质通道将 Na^+ 泵出细胞，哺乳动物细胞就是通过这种方式维持平衡状态并防止 Na^+ 在胞内不断聚集的。由于这种 Na^+ 主动泵出细胞的速率等于它从其他通道进入细胞的速率，因此看起来就像 Na^+ 不能穿过细胞膜一样。其实，哺乳动物细胞处于一种稳定状态，而并非平衡状态，细胞必须消耗 ATP 形式的能量以阻止离子的净跨膜流动。

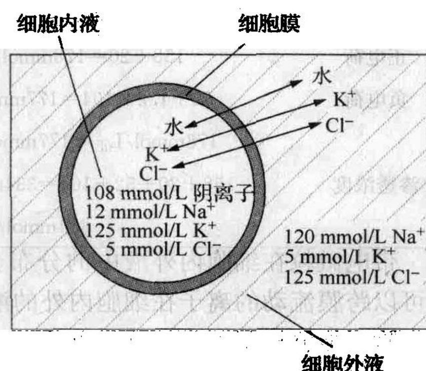


图 3.7　一种细胞模型，其细胞内液和外液中各种离子的浓度与普通哺乳动物细胞相似，胞外溶液的体积远大于细胞体积（图中没有将其全部表示出来）

例 3.4　简化细胞模型如图 3.7 所示，假设各种离子浓度如下所列。请问该细胞是否处于平衡状态，为什么？

离子	胞内浓度（mmol/L）	胞外浓度（mmol/L）
K^+	158	4
Na^+	20	163
Cl^-	52	167
A^-	104	—

解：细胞是处于平衡状态。由于细胞内的正负电荷量相等，细胞外的正负电荷量也相等，并且胞内渗透浓度等于胞外渗透浓度，因此细胞既处于电荷平衡也处于渗透平衡。具体计算如下：

	胞内	胞外
正电荷	158＋20＝178mmol/L	4＋163＝167mmol/L
负电荷	52＋1.2×104＝177mmol/L	167mmol/L
	$178mmol/L_{正} \approx 177mmol/L_{负}$	$167mmol/L_{正} = 167mmol/L_{负}$
渗透浓度	158＋20＋52＋104＝334mmol/L	4＋163＋167＝334mmol/L
	$334mmol/L_{胞内} = 334mmol/L_{胞外}$	

■

带电微粒在细胞内外液中的分布会导致跨膜电位的产生，电位的大小取决于各种可以跨膜流动的离子在细胞内外的浓度，本书第 11 章将详细讨论这个问题。

除了能够控制细胞的体积以外，细胞膜也是大分子等其他物质进出细胞的徐径。各种物质可以通过胞吞作用进入细胞（见图 3.8a），也可以通过胞吐作用离开细胞（见图 3.8b）。胞吞时，细菌等细胞外的物质被小片质膜包裹，形成小囊泡；然后，小囊泡与细胞膜融合，被吞并，囊泡内容物就流入细胞内。胞吐时，细胞内的物质被质膜包裹，也形成小囊泡，然后，小囊泡移向细胞膜，并与细胞膜融合，将其内容物释放到细胞外。

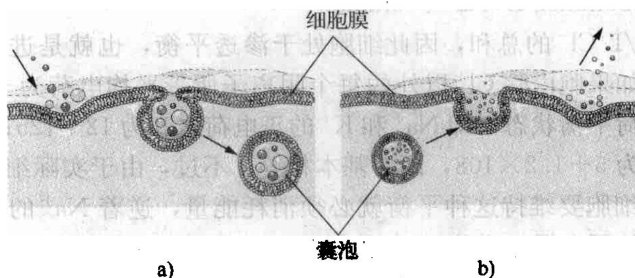


图 3.8 某些物质由于体积太大，不能穿过细胞膜上的蛋白质通道，但可以通过胞吞（图 a）和胞吐（图 b）方式进出细胞

3.2.2 细胞质和细胞器

细胞质包含了细胞液和细胞器，细胞液中含有 Na^+、K^+、Cl^- 等离子和葡萄糖等分子。细胞器有两类：一类被质膜包裹，有细胞核、粗面内质网、光面内质网、

高尔基体、溶酶体和线粒体；另一类无质膜包裹，有核仁、核糖体、中心粒、微绒毛、纤毛、鞭毛以及细胞骨架中的微管、中间丝和微丝等。

如图 3.4 所示，细胞核由核膜（也是双层膜）和核质组成，核质中含有各种离子、酶、核苷酸、蛋白质、DNA 和少量 RNA。细胞核的 DNA 中含有指导生命过程的指令信息。核膜上的核孔是蛋白质通道，允许各种离子和 RNA 通过，但蛋白质和 DNA 一般不能通过。大多数细胞核含有一个或者多个核仁，每个核仁都包含有 DNA、RNA 和蛋白质，并可以合成核糖体的各种成分，供细胞生产蛋白质。

光面内质网和粗面内质网、高尔基体以及各种囊泡（见图 3.4、3.9a 和 3.9b）组成了细胞的质膜系统，用于输送构建质膜所需的蛋白质和脂类，并积累和存储某些特殊用途的蛋白质和脂类。内质网还是钙离子的存储部位。粗面内质网与光面内质网的不同之处是前者外表面粘附有核糖体，核糖体是蛋白质合成的场所。粗面内质网外表面上合成的蛋白质通过内质网膜进入内质网内侧，并在那里被加上非蛋白质侧链，然后这些经过修饰的蛋白质移向光面内质网，并在那里被包装成囊泡。光面内质网还生产脂类并用于生成囊泡，它也负责释放存储的钙离子。光面内质网生成的囊泡释放后会粘附到高尔基体上，在此，囊泡内容物被释放、修饰并重新包装成新的囊泡。其中有一类囊泡就是溶酶体，内含消化酶，可降解那些通过胞吞作用进入细胞的物质。其他囊泡含有激素和神经递质等蛋白质，通过胞吐作用由细胞分泌到胞外。

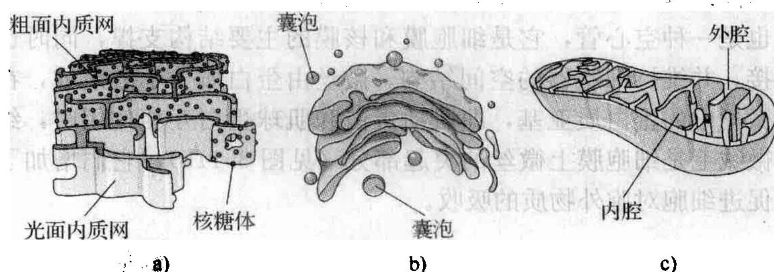


图 3.9　几种细胞器

注：内质网（图 a）、高尔基体和囊泡（图 b）组成了细胞的质膜系统。内质网上的小颗粒是核糖体。附有核糖体的内质网称为粗面内质网，没有核糖体的内质网称为光面内质网。线粒体（图 c）具有两层质膜，其中内膜将线粒体内分成内腔和外腔两个部分，各自含有不同浓度的酶、底物和氢离子。内腔与外腔之间的电压差和化学浓度梯度为生产 ATP 提供了所需的能量。

线粒体（见图 3.9c 和图 3.10）含有两层质膜：外膜包裹整个细胞器，内膜则将线粒体内隔成了两个部分。细胞所需的 ATP 中大约有 95%是由线粒体通过一系列需氧反应产生的，该反应还同时产生二氧化碳。与其他细胞器不同的是线粒体内含有自身 DNA。大多数有性繁殖生物体，如人类，其线粒体直接来源于母体的卵细胞，因为父体的精子只能为后代提供半套染色体的 DNA。

微管、中间丝和微丝为细胞提供了结构支持并协助细胞运动。微管由微管蛋白的螺旋亚基组成，是一种中空的长圆柱形结构。在细胞分裂过程中，细胞质中靠近

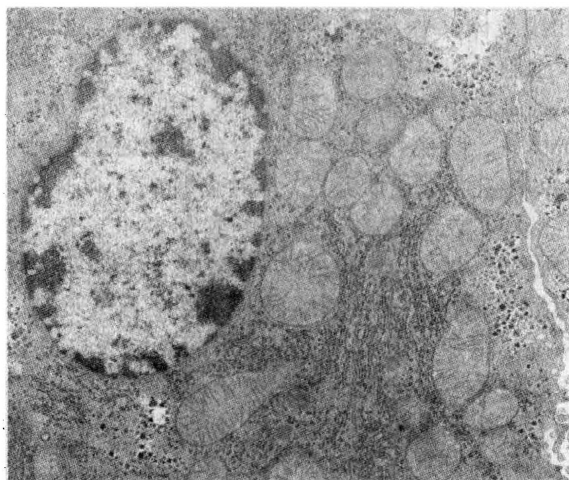


图 3.10 正常小鼠肝细胞放大 8000 倍的电镜扫描图像

注：左边较大的圆形细胞器是细胞核，其他较小的圆形或椭圆形细胞器是线粒体，它们具有不同的切片角度。平行排列的细长的膜结构是内质网，内质网上的小颗粒就是核糖体。（此照片由 North Carolina State University 电镜中心的 Valerie Knowlton 馈赠。）

细胞核的一个特殊区域（即中心体）含有两个中心粒，这对中心粒相互垂直，每个中心粒就是由围绕其中心辐射状排列的 9 组微管三联体构成，形状就像车轮上的辐条。中间丝也是一种空心管，它是细胞膜和核膜的主要结构支撑，同时也辅助细胞之间形成连接，并维持细胞器的空间结构。微丝由蛋白质分子链组成，在一般细胞中都有。当微丝中的蛋白质亚基，即肌动蛋白和肌球蛋白互相作用时，细胞就可以发生移动。微绒毛是细胞膜上微丝的突起部分（见图 3.11b)，它们增加了细胞的表面积，可以促进细胞对胞外物质的吸收。

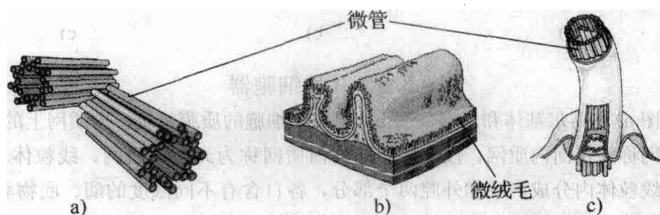


图 3.11 细胞的中心体含有 2 颗中心粒（图 a)，两者之间形成直角。这种中心粒由微管组成，它们在细胞分裂过程中用于固定牵引细胞遗传物质平分的微管，因此具有非常重要的作用。微绒毛（图 b）是细胞膜的突起部分，整齐排列的绒毛在小肠粘膜上就像微小的手指状突起，有助于增加细胞吸收营养物质的表面面积。纤毛（图 c）排列在呼吸道中，其摆动作用有助于将混合在粘液中的细菌和微粒排出肺外

纤毛和鞭毛是细胞的骨架成分（见图 3.11c)，都是杆状结构，由外周的 9 对微管和中间的 2 根单微管组成，这些杆状结构再锚定在与中心粒结构相似的基座上。

鞭毛的作用像一根尾巴，可以推动精子等细胞前行。纤毛通常比鞭毛短，但数量比鞭毛多，如呼吸道管壁上排列的那种特异细胞就具有纤毛，纤毛的摆动作用有助于将混合在粘液中的细菌和微粒排出肺外。

3.2.3　DNA 以及基因表达

DNA（见图 3.3）存在于细胞核和真核生物的线粒体中。在有性繁殖的生物体中，细胞核 DNA 包含了来自父母双方的遗传信息；而线粒体 DNA 只来源于母体。细胞核 DNA 缠绕在蛋白质柱（即核小体）上，形成染色体对。如图 3.12 所示，人类含有 22 对常染色体和 1 对性染色体，性染色体 XX 表示女性，XY 则表示男性。如果将一个人体细胞（除精子和卵子外的任何细胞）中的所有 46 条染色体上的 DNA 头尾相连，那么，可以形成一条宽 2nm、长 2m 的 DNA 带。每个染色体上都含有成千上万个独立的基因，它们是遗传性状的信息单位。每个基因都位于特定染色体上的特定位置，并且编码 3 种 RNA（即核糖体 RNA、信使 RNA、转运 RNA）中的一种。人类基因组计划始于 1990 年，其目标是确定人类整个半套染色体（23 对染色体中每对取一条）上 3000 多个基因的位置，并测定它们的核苷酸序列，共约有 300 万对核苷酸！有关人类基因组计划的详细内容请参阅第 13 章。

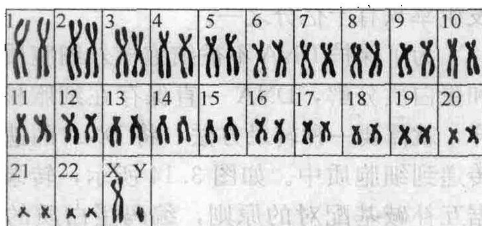


图 3.12　正常男性的染色体组型

注：图中按照从大到小的顺序显示了 22 对常染色体，以及 XY 性染色体。

如图 3.13 所示，细胞分裂时发生 DNA 复制，这是一种半保留复制。首先，酶

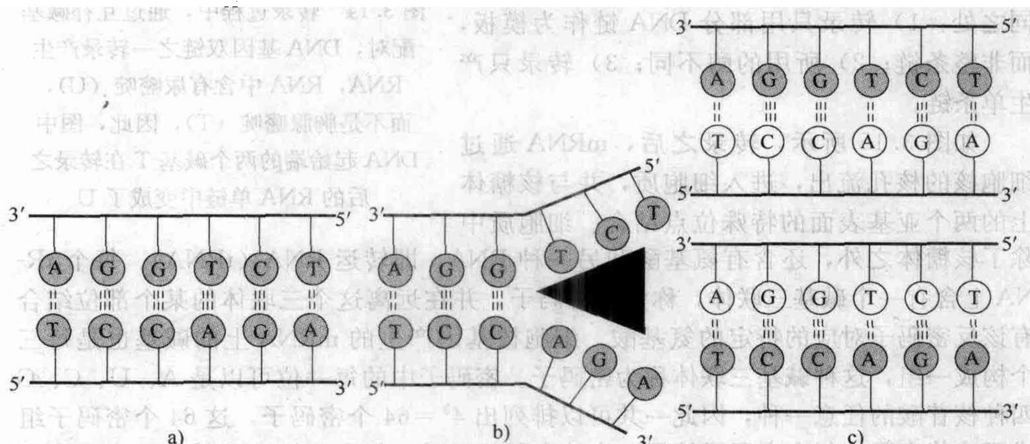


图 3.13　复制过程中，DNA 双螺旋（图 a）首先由解旋酶解链（图 b 中的黑三角区域所示），再由另一种酶——DNA 聚合酶，按照从 5′端到 3′端的方向，分别复制解链后的每一条单链。其中 5′到 3′方向的那条链可以连续复制，而 3′到 5′方向的反向链则需要先合成 5′到 3′方向的各个小片段，再通过 DNA 连接酶将它们连接起来。复制结束时产生两条相同的 DNA 双链

将双螺旋解链，然后互补碱基输送到解链的核苷酸链旁，并相互连接形成互补的DNA单链。完成复制之后，每条解链的DNA被复制成两条，于是就形成了两条DNA双螺旋链，每条双螺旋中都包含了一条原DNA链和一条新合成的DNA链。这样，每个子细胞从分裂的母细胞中都获得相同的遗传信息。在复制过程中，某些酶负责检查复制的准确性，还有一些酶负责修正碱基配对的错误，从而使复制错误的发生率只有十亿分之一。

为了保护DNA不被细胞核外细胞质中的酶和蛋白质分解，DNA一直保存在细胞核内，这样，就需要一种转录方法，将DNA的遗传信息传递到细胞质中。如图3.14所示，转录时，根据互补碱基配对的原则，编码蛋白质的基因的核苷酸序列被转录成信使RNA（即mRNA）。例如，某段DNA序列为TACGCTCCGATA，则其mRNA序列应该是AUGCGAGGCUAU。实际转录过程还要复杂一些，直接由含有内含子的DNA转录生成mRNA，再经过剪切，才能产生最终的mRNA。这种最终的mRNA还有一条尾巴，称为聚A尾，由100～200个腺苷酸（A）连接而成；并且在它的另一端连有一个由核苷酸组成的帽子结构，该核苷酸上结合了一个甲基和几个磷酸基团。转录与复制有如下不同之处：1）转录只用部分DNA链作为模板，而非整条链；2）所用的酶不同；3）转录只产生单条链。

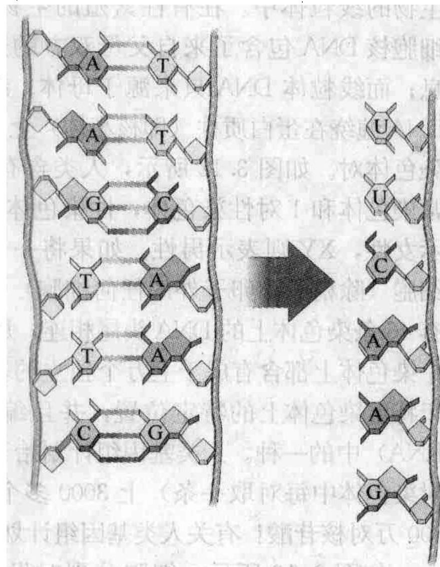


图3.14 转录过程中，通过互补碱基配对，DNA基因双链之一转录产生RNA，RNA中含有尿嘧啶（U），而不是胸腺嘧啶（T），因此，图中DNA起始端的两个碱基T在转录之后的RNA单链中变成了U

如图3.15所示，转录之后，mRNA通过细胞核的核孔流出，进入细胞质，并与核糖体上的两个亚基表面的特殊位点结合。细胞质中除了核糖体之外，还含有氨基酸和另一种RNA，即转运RNA（tRNA）。每个tRNA上含有一个碱基三联体，称为反密码子，并在远离这个三联体的某个部位结合有该反密码子对应的特定的氨基酸。细胞核基因产生的mRNA上的碱基也是每三个构成一组，这种碱基三联体称为密码子。密码子中的每一位可以是A、U、C、G四种核苷酸的任意一种，因此一共可以排列出$4^3=64$个密码子，这64个密码子组成了遗传密码，每个密码子编码一个特定的氨基酸，但有些氨基酸对应于多个密码子（见表3.1）。例如，AUG是甲硫氨酸惟一的mRNA密码子，它通常指示翻译的起始位置㊀，而亮氨酸的密码子却有UUA、UUG、CUU、CUC、CUA和CUG。因

㊀ 根据遗传信息生成蛋白质的过程称为翻译。——作者注

此，将甲硫氨酸运送到核糖体的tRNA上的反密码子为UAC，而运送亮氨酸的tRNA上的反密码子可以是AAU、AAC、GAA、GAG、GAU和GAC。

翻译过程中，mRNA与核糖体相结合，tRNA则根据mRNA上的密码子不断地将氨基酸运输到正在生长的肽链上，每个新转运来的氨基酸都与前一个氨基酸之间形成肽键连接。一旦氨基酸结合到增长的肽链上之后，它就与tRNA脱离。该tRNA重新回到细胞质中并根据其反密码子结合新的氨基酸分子。这种翻译过程一直持续到遇到mRNA上的终止密码子（UAA、UAG或者UGA）为止。然后，蛋白质脱离核糖体，进入细胞质，或者进入粗面内质网中完成进一步的修饰。

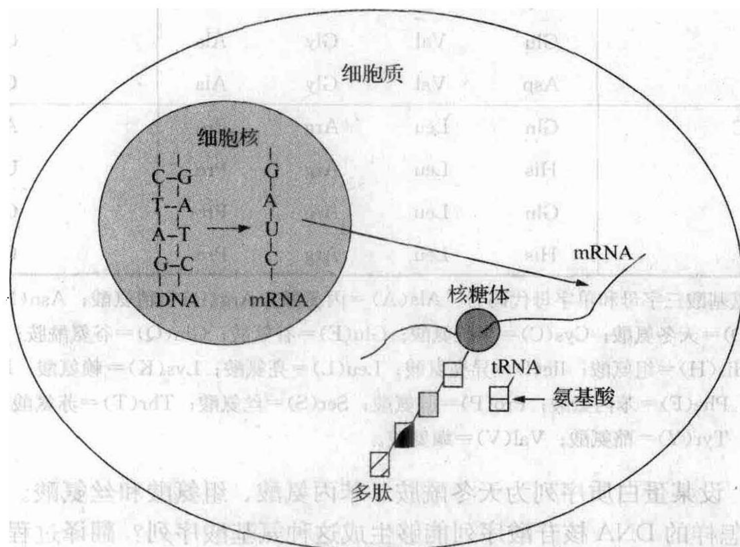


图3.15 在细胞核内完成DNA转录和处理之后，mRNA从细胞核转移到细胞质。在细胞质中，mRNA与核糖体结合，并开始蛋白质翻译过程。在翻译过程中，tRNA将氨基酸转运到不断增长的肽链上，转运哪个氨基酸则由mRNA上的三联碱基密码子决定，mRNA的每个密码子与特定的tRNA上的反密码子互补。每个tRNA都有一个特定的氨基酸结合在与反密码子相对的位点上。例如，mRNA上的密码子CUG与携带亮氨酸的tRNA上的反密码子GAC互补，于是就将亮氨酸添加到肽链中

表3.1 遗传密码

第一位碱基	第二位碱基				第三位碱基
	A	U	G	C	
A	Lys	Ile	Arg	Thr	A
	Asn	Ile	Ser	Thr	U
	Lys	Met-Start	Arg	Thr	G
	Asn	Ile	Ser	Thr	C

（续）

第一位碱基	第二位碱基				第三位碱基
	A	U	G	C	
U	Stop	Leu	Stop	Ser	A
	Tyr	Phe	Cys	Ser	U
	Stop	Leu	Trp	Ser	G
	Tyr	Phe	Cys	Ser	C
G	Glu	Val	Gly	Ala	A
	Asp	Val	Gly	Ala	U
	Glu	Val	Gly	Ala	G
	Asp	Val	Gly	Ala	C
C	Gln	Leu	Arg	Pro	A
	His	Leu	Arg	Pro	U
	Gln	Leu	Arg	Pro	G
	His	Leu	Arg	Pro	C

注：表中氨基酸三字母和单字母代码为：Ala(A)＝丙氨酸；Arg(R)＝精氨酸；Asn(N)＝天冬酰胺；Asp(D)＝天冬氨酸；Cys(C)＝半胱氨酸；Glu(E)＝谷氨酸；Gln(Q)＝谷氨酰胺；Gly(G)＝甘氨酸；His(H)＝组氨酸；Ile(I)＝异亮氨酸；Leu(L)＝亮氨酸；Lys(K)＝赖氨酸；Met(M)＝甲硫氨酸；Phe(F)＝苯丙氨酸；Pro(P)＝脯氨酸；Ser(S)＝丝氨酸；Thr(T)＝苏氨酸；Trp(W)＝色氨酸；Tyr(Y)＝酪氨酸；Val(V)＝缬氨酸。

例 3.5 设某蛋白质序列为天冬酰胺、苯丙氨酸、组氨酸和丝氨酸。如果不存在内含子，则怎样的 DNA 核苷酸序列能够生成这种氨基酸序列？翻译过程中负责转运这些氨基酸的 tRNA 反密码子又是哪些？

解：表 3.1 的遗传密码列出了对应于各种氨基酸的 mRNA 密码子序列，根据碱基互补配对的原则，可以利用这些 mRNA 密码子序列确定原始 DNA 序列和 tRNA 的反密码子序列。注意，DNA 用的是胸腺嘧啶 T，而不是尿嘧啶 U，而 mRNA 和 tRNA 都用 U，而不是 T。碱基配对的例子请参见图 3.3 和图 3.14。本题答案如下：

	天冬酰胺 (Asn)	苯丙氨酸 (Phe)	组氨酸 (His)	色氨酸 (Ser)
mRNA 密码子	AAU 或 AAC	UUU 或 UUC	CAU 或 CAC	UC（A，G，U 或 C）
DNA	TTA 或 TTG	AAA 或 AAG	GTA 或 GTG	AG（T，C，A 或 G）
tRNA 反密码子	UUA 或 UUG	AAA 或 AAG	GUA 或 CUG	AG（U，C，A，或 G）



3.3 组织

如图 3.16 所示，组织是由多个细胞群及其周围物质构成并具有一种或多种特定功能的生物体。人体中有 4 种主要的组织类型：上皮组织（epithelial tissue）、结缔组织（connective tissue）、肌肉组织（muscle tissue）和神经组织（nervous tissue）。上皮组织中细胞要么形成单层或多层的片状结构，要么组成具有分泌功能的腺体。其特点是具有游离的表面，例如肠道内表面、皮肤外表面等；还具有基底膜。上皮组织的主要功能包括吸收功能（如小肠）、分泌功能（如腺体）、运输功能（如肾小管）、排泄功能（如汗腺）、保护功能（如皮肤）以及感觉器接受功能（如味蕾）。结缔组织最丰富，分布最广泛。它有不同的种类：有疏松结缔组织，如器官周围和器官之间的疏松的交织纤维；有不规则的致密结缔组织，如包绕各种器官的保护性被膜；还有规则的致密结缔组织，如韧带和肌腱。特异性的结缔组织有血液（见图 3.16b）、骨组织、软骨和脂肪组织等。肌肉组织通过其特

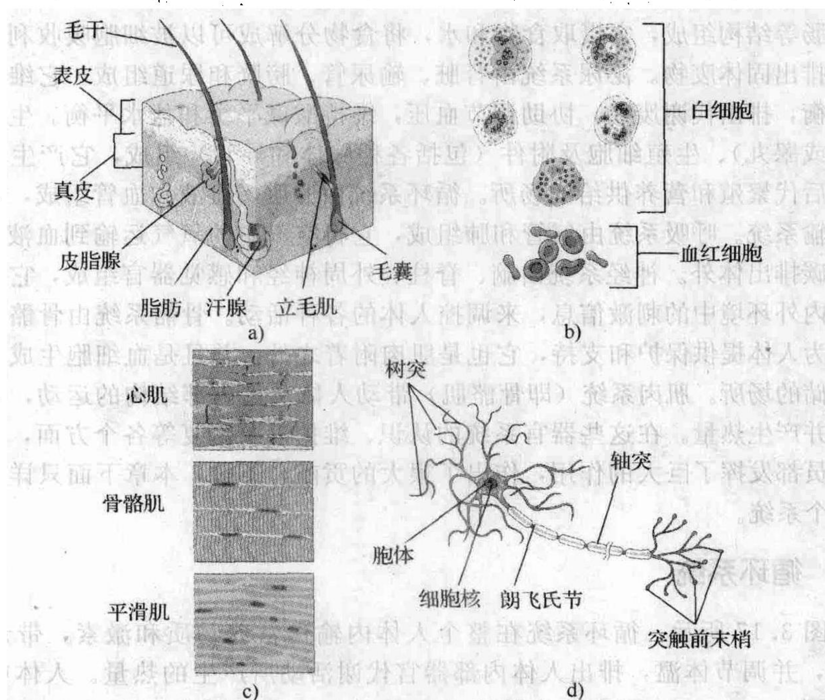


图 3.16　4 种组织

注：皮肤（图 a）是一种上皮组织，它有助于保护身体；血液（图 b）是一种特殊的结缔组织；肌肉组织（图 c）有 3 种不同的类型：心肌、骨骼肌和平滑肌；运动神经元（图 d）是一种神经组织，它能将中枢神经系统的电脉冲信号传递给肌肉等效应器。

异化的细胞产生身体的运动，肌肉细胞受到刺激后可以收缩，然后再回复到原有的舒张状态。图 3.16c 显示了 3 种不同的肌肉组织：附着于骨骼上的骨骼肌、组成血管壁的平滑肌、只存在于心脏的心肌。神经组织由神经元和神经胶质细胞组成（见图 3.16d），其中神经元具有传导电脉冲信号的功能，而胶质细胞则具有保护、支持和营养神经元的功能。

3.4 人体主要器官系统

几种组织结合在一起所形成的能够完成复杂功能的结构称为器官，共同执行特定任务的多个器官则组成了器官系统。人体共有 11 个主要器官系统：表皮系统、内分泌系统、淋巴系统、消化系统、泌尿系统、生殖系统、循环系统、呼吸系统、神经系统、骨骼系统和肌肉系统。表皮系统由皮肤、毛发、指甲和各种腺体等组成，它为人体提供保护。内分泌系统包括甲状腺和肾上腺等无管腺体，它们分泌激素，用于调控细胞内的各种化学反应。淋巴系统包括腺体、淋巴结、淋巴和淋巴管，它将多余的体液和蛋白质回收到血液中，并协助人体预防感染和组织损伤。消化系统由胃、肠等结构组成，它摄取食物和水，将食物分解成可以被细胞吸收利用的小分子，并排出固体废物。泌尿系统由肾脏、输尿管、膀胱和尿道组成，它维持人体的体液平衡，排出代谢废物，协助调节血压，维持酸碱平衡和盐水平衡。生殖系统由卵巢（或睾丸）、生殖细胞及附件（包括各种腺体和导管）组成，它产生卵子或精子，是后代繁殖和营养供给的场所。循环系统由心脏、血液和血管组成，它是人体内的运输系统。呼吸系统由气管和肺组成，它将空气中的氧气运输到血液中，并将二氧化碳排出体外。神经系统由脑、脊柱、外周神经和感觉器官组成，它通过感知并响应内外环境中的刺激信息，来调控人体的各种活动。骨骼系统由骨骼和软骨组成，它为人体提供保护和支持，它也是肌肉附着之处，并且是血细胞生成以及钙和磷酸存储的场所。肌肉系统（即骨骼肌）带动人体及其内部结构的运动，保持人体姿势，并产生热量。在这些器官系统的认识、维护以及修复等各个方面，生物医学工程人员都发挥了巨大的作用，作出了很大的贡献。不过，本章下面只详细介绍上述后 5 个系统。

3.4.1 循环系统

如图 3.17 所示，循环系统在整个人体内输送营养物质和激素，带走组织中的废物，并调节体温，排出人体内部器官代谢活动所产生的热量。人体中每个活细胞周围 10～100μm 距离之内就有毛细血管，毛细血管是很细小的血管，其血管壁只有一个细胞的厚度，直径只有 8μm，相当于一个血红细胞的大小。细胞与毛细血管之间如此接近，使得氧气、二氧化碳等多数小分子溶质都能够在浓度梯度和压力梯度作用下扩散运输，要么从细胞进入毛细血管，要么从毛细血管进入细胞。

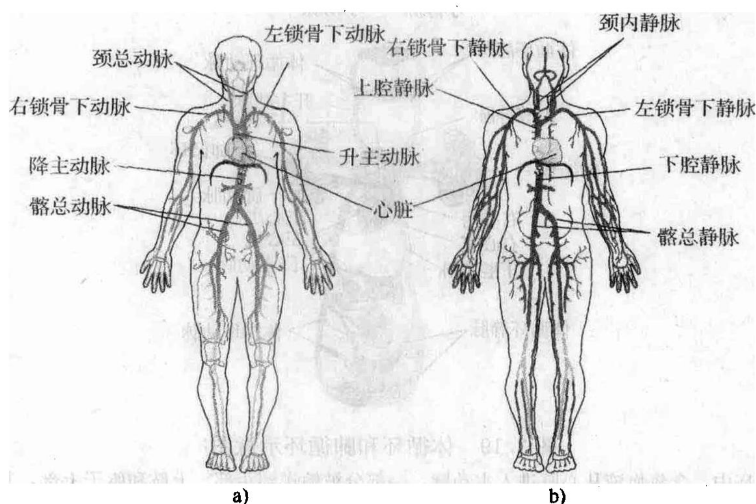


图 3.17 图 a 是人体主要动脉分布图，动脉将血液从心脏运输到全身；
图 b 是人体主要静脉分布图，静脉将血液运送回心脏

如图 3.18 所示，心脏就像泵站，由左侧和右侧两个泵组成。每个泵都由一个接收血液的心房和一个将血液泵出心脏的心室组成。右侧的泵接收经全身循环而来的富含二氧化碳的缺氧静脉血，并将其泵入肺部；左侧的泵则接收经肺去除大部分二氧化碳的富氧动脉血，并将其泵入全身。如图 3.19 所示，由心脏通向肺和由肺返回心脏的血管系统组成了肺循环；由心脏通向全身其他各个组织和由这些组织返回心脏的血管系统则组成了体循环。将血液从心脏运输到全身各处的血管称为动脉，而将血液输送回心脏的血管称为静脉。含氧少的血液称为静脉血，含氧多的血液称为动脉血。肺动脉是惟一输送静脉血的动脉，而肺静脉是惟一输送动脉血的静脉。成人体内平均含有 5

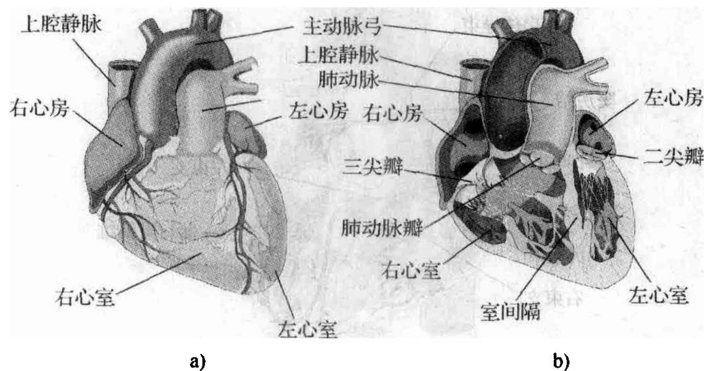


图 3.18 图 a 是心脏外表的前视图；图 b 是切除外表面之后的心脏前视图，
其中可见 4 个内部腔室——左、右心房和左、右心室，以及几个瓣膜

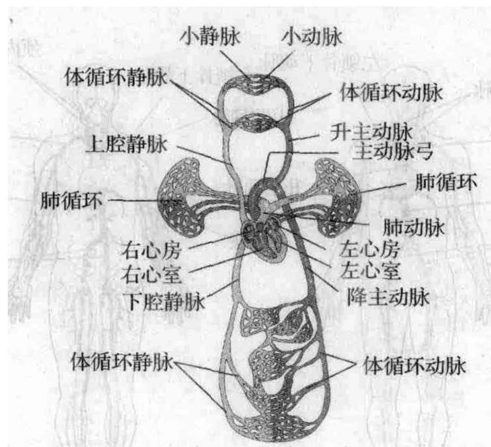


图 3.19 体循环和肺循环示意图

注：体循环中，含氧血液从心脏进入主动脉，一部分被输送到头部、上肢和躯干上部，其余部分则被输送到躯干下部和下肢。血液从主动脉进入其他动脉血管，流经小动脉，进入毛细血管网，在这里血液与其周围细胞交换营养物质、激素、氧气、二氧化碳和废物等。然后，血液从毛细血管网经小静脉进入静脉。身体上部的血液经上腔静脉回到右心房，身体下部的血液则经下腔静脉回到右心房。肺循环中，右心房的血液通过右心室经肺动脉进入肺循环系统，在肺的毛细血管中经过气体交换之后，重新变成含氧血液，并通过肺静脉回到左心房。最后，血液从左心房流入左心室，通过主动脉进入体循环，开始新一轮循环。

升血液，其中约有 80%～90%的血量分布在体循环中；体循环的 75%血量分布在静脉中，20%在动脉中，还有 5%则在毛细血管中。心输出量（cardiac output）等于心率和心脏每搏泵血量（即每搏输出量（stroke volume））的乘积。心脏每搏输出量约为 80mL，因此，心脏大约需要 60 次搏动才能使一个血红细胞走完整个体循环。

对于正常心脏，心动周期（cardiac cycle）是指心脏各房室一次收缩和一次舒张构成的活动周期，它起始于心脏窦房结起搏细胞产生的电脉冲（见图 3.20）。细胞的

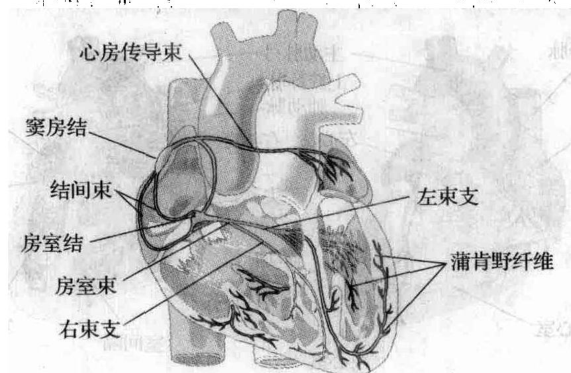


图 3.20 心脏动作电位的传导

注：窦房结的自律性起搏细胞首先去极化，然后其动作电位传过心房，当兴奋波传导到房室结时，其传播速度减慢，之后又快速通过房室束和蒲肯野纤维，最后到达心室细胞。

这种快速电脉冲变化是各种离子跨膜运动的结果。在此过程中，由于细胞膜对 Na^+ 离子的通透性发生了巨大的变化，胞外 Na^+ 离子涌入细胞内，于是，跨膜电位由原来的胞内比胞外负（约为－90mV）变成了胞内比胞外正（约为 20mV），这种电位变化称为去极化。短时间（<0.3s）之后，膜的其他变化和钠钾离子泵的作用使细胞膜复极化，重新回到原来的平衡状态。这种跨膜电位极性快速逆转后又重新恢复的整个电变化过程称为动作电位。普通成人休息状态下窦房结平均每 0.83s 去极化一次，相应的静息心率是每分钟 72 次，大约每个心动周期的 5/8 时间是心脏舒张期，其余 3/8 时间是心脏收缩期。

心肌细胞相互之间紧密耦联，因此动作电位可以沿着一个个细胞传播开来。下面介绍一个心动周期中动作电位兴奋的传播过程。兴奋波大约以 1m/s 的速度传过心房，使心肌细胞去极化，心房收缩。心房收缩时将血液从心房挤入心室，即右心房到右心室，左心房到左心室（见图 3.21）。当兴奋波传到房室结时，其传播速度降至约 0.05m/s，使得心室具有足够的时间充盈血液。然后，兴奋波从房室结传导到一种特殊的传导组织——浦肯野纤维，并以极快的速度（约 3m/s）迅速将兴奋波传播到两侧心室。兴奋波到达心室后，在心室组织中传播的速度大约为 0.5m/s。这样，两个心室就同时收缩，将血液从右心室泵入肺动脉，从左心室泵入主动脉。

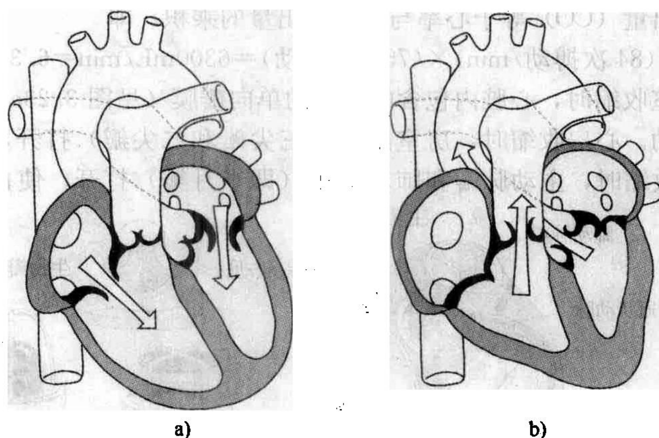


图 3.21 心动周期中动作电位兴奋的传播过程

注：图 a 为心动周期的第一阶段，心房收缩，并将血液挤入心室；图 b 为心动周期的第二阶段，心房舒张，心室收缩，并将血液泵入肺部（即进入肺循环）或者泵入身体的其他部分（即进入体循环）。

如图 3.22 所示，心电图（electrocardiogram，ECG）是测量上述心脏整个电位变化的一种方法。其中，P 波表示心房的去极化，QRS 波表示心室的去极化，T 波表示心室的复极化，心房的复极化则被心室的去极化掩盖了。ECG 各个部分幅值和持续时间的变化是医生诊断的重要信息。有很多生物医学工程人员致力于 ECG 记录和分析方法的研究。

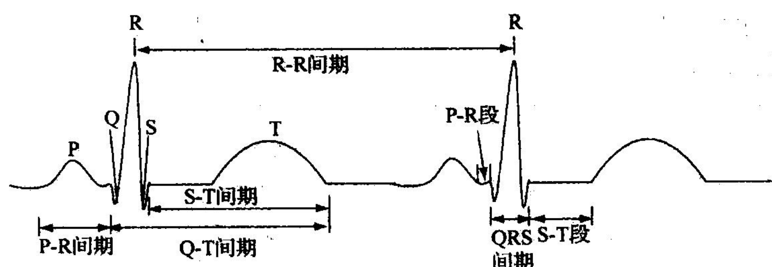


图 3.22 常规Ⅱ导联心电图

注：将正电极放在左腿，负电极放在右臂，就可以测到这种人体体表的普通 ECG 波形。垂直方向表示电压幅值，水平方向表示时间。从图中很容易辨别 P、R、T 波，这些波是心脏不同部位细胞离子活动的结果。各个不同区段的持续时间和波形包含了心脏及其传导系统的健康信息。其中 R-R 间期可用于计算心率。

例 3.6 假设 6.4s 内 ECG 中出现 10 个 R 波，则其心率是多少？

解：10 个 R 波之间包含有 9 个 R-R 间期（见图 3.22），也就是 9 次心脏搏动，因此心率为

$$(9\text{ 次搏动}/6.4\text{s})\times(60\text{s/min})=84\text{ 次搏动/min}$$ ■

例 3.7 如果每搏输出量为 75mL，则例 3.6 所述心脏的心输出量（CO）为多少？

解：心输出量（CO）等于心率与每搏输出量的乘积，即

$$\text{CO}=(84\text{ 次搏动/min})\times(75\text{mL/搏动})=6300\text{mL/min}=6.3\text{L/min}$$ ■

心房和心室收缩时，心脏内包含的特殊的单向瓣膜（见图 3.23a）使血液只能沿正确的方向流动。心房收缩时，房室瓣（即三尖瓣和二尖瓣）打开，血液由心房流向心室。心室收缩时，主动脉瓣和肺动脉瓣（即半月瓣）打开，使血液从心脏泵入

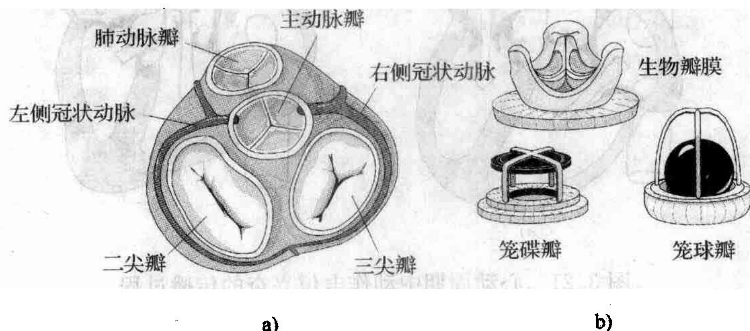


图 3.23 图 a 为房室瓣（即三尖瓣和二尖瓣）以及肺动脉瓣和主动脉瓣（即半月瓣）；图 b 为三种人工瓣膜：生物瓣膜、笼碟瓣和笼球瓣⊖，它们可以用于替代发生病变或畸形的心脏瓣膜

⊖ 生物瓣膜是指取自动物或人的瓣膜，笼碟瓣和笼球瓣属于机械瓣膜。——译者注

动脉；同时房室瓣关闭以阻止血液从心室倒流回心房。主动脉瓣和肺动脉瓣还分别阻止动脉血液倒流回左右心室。如果心脏的瓣膜发生钙化、病变或者在胚胎发育期间产生畸变，那么可以实施手术，用人工瓣膜（见图 3.23b）来替代，人工瓣膜是生物医学工程人员与医生合作开发的成果。

血压可以通过直接或间接（无创）的方法测量。直接测量方法需要将与压力传感器相连的导管或针插入到静脉或动脉中。间接测量的第一种常用方法是用血压计测量（sphygmomanometry），也就是用袖带裹住前臂，并将袖带充气，给动脉施加足够大的压力，阻止血流通过动脉，然后再慢慢地释放袖带中的空气，同时用听诊器监听血流的声音变化。当袖带压力降到刚刚低于收缩压（即心室收缩射血时达到的最高血压）时，开始听到科罗特科夫（Korotkoff）音，这表示听诊器下的动脉开始有血液冲过。当袖带压力降到低于舒张压（即心室舒张末期最低血压）时，科罗特科夫音变得低沉，并逐渐消失。第二种间接测量方法是示波法（oscillometric method），它利用微处理器控制袖带反复进行周期性的充气和缓慢放气。每当血液在袖带施加的阻力下冲过血管时，血流形成的湍流会引起动脉血管壁的轻微振动。振动刚开始出现时的血压就是收缩压，随后，振动的幅值逐渐减小，幅值突然快速下降的那一刻就对应于舒张压。第三种间接测量法是超声法，它根据血流中血红细胞对声波反射所产生的多普勒频移来测定血流速度，从而测量血压。

体循环中，正常成人血液离开心脏进入主动脉时的平均血压约为 100mmHg，收缩压约为 120mmHg，舒张压约为 80mmHg。然后，血液流入中等大小的动脉和小动脉（见图 3.17a），再由小动脉通向毛细血管，毛细血管中的平均血压约为 30 mmHg。随后血液进入小静脉，由小静脉流经中等大小静脉，再进入大静脉，最后汇入腔静脉，腔静脉的平均血压只有 10mmHg。腔静脉的血液流入右心房返回心脏。肺循环（见图 3.19）中的血液由肺动脉进入小动脉，然后进入肺内的毛细血管网，最后经左心房回到心脏。大动脉和大静脉中的血流速度最快，主动脉的血流速度约为 30～40cm/s，腔静脉的血流速度约为 5cm/s；而毛细血管网中的血流速度最慢，只有 1mm/s，它是营养物质、代谢废物、气体以及激素的交换场所。肺循环的血压比体循环的血压低，其平均收缩压约为 25mmHg，舒张压约为 10mmHg。这是因为右心室比左心室小，其泵血力量也较弱，并且肺血管的阻力也较小。

例 3.8 测得某人的血压为 118mmHg/79mmHg（即收缩压/舒张压），请问他的脉搏压和平均动脉压分别是多少？

解： 脉搏压（pulse pressure）等于收缩压（118mmHg）与舒张压（79mmHg）之差，本题应该是 39mmHg。

平均动脉压（mean arterial pressure）是指动脉中的平均血压，通常用舒张压加上 1/3 脉搏压来计算，因此，本题的平均动脉压为 92mmHg。 ■

3.4.2 呼吸系统

如图 3.24a 所示，呼吸系统在外界空气和体内气体交换界面之间形成运输通道，

体内气体交换界面是循环系统的血液与空气之间进行扩散交换的场所。呼吸系统分为导管区（conduction zone）和呼吸区（respiratory zone）两大部分。导管区包括口、鼻、鼻窦、咽、气管、支气管和细支气管，当空气通过这些导管进入体内时，空气被加热、湿润、过滤和净化。导管区细胞会分泌粘液，可以吸附空气中大于 6μm 的小颗粒，阻止它们进入呼吸区。排列在气管和支气管表面的上皮细胞具有纤毛，纤毛协调一致的摆动可以将粘液运送到咽部，然后被吞咽或者咳出。呼吸区由带有肺泡突起的呼吸性细支气管和肺泡囊末端组成，它是空气与血液之间进行气体交换的场所（见图 3.24b）。呼吸区构成了肺的绝大部分重量。

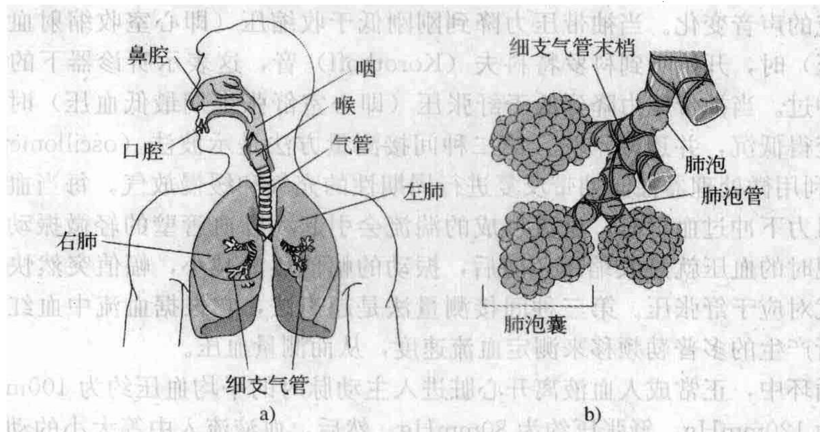


图 3.24 a）呼吸系统由人体中空气进出的各种通道组成 b）肺内的细支气管末梢和肺泡囊含有许多肺泡，肺内气体与毛细血管血液之间的气体交换就在这些部位完成

肺的特性可以用顺应性、弹性和表面张力等物理量来描述。顺应性（compliance）是指肺在力的作用下的可扩张程度。正常肺的可扩张性是玩具气球的 100 倍。弹性（elasticity）是指肺和其他胸腔结构从扩张状态恢复其原始大小的能力，弹性有助于呼气阶段气体从肺中排出。表面张力（surface tension）由肺泡中的液体薄膜产生，可以抵抗扩张。它产生一种内向的力，施加在肺泡上。根据 Laplace 定律，肺泡内压力的大小与表面张力成正比，与肺泡的半径成反比。因此，半径小的肺泡内的压力要比相邻的半径大的肺泡内的压力大，从而导致小肺泡中的气体进入大肺泡，这会使小肺泡塌陷。但是，正常的肺不会发生这种情况，因为肺泡液体中含有磷脂，它具有表面活性剂的作用。表面活性剂可以降低肺泡的表面张力，使肺泡在呼气阶段变得比较小但并不塌陷。未发育成熟的婴儿由于体内缺乏肺表面活性剂，经常会患呼吸窘迫综合症，这时，可以借助机械呼吸机或者表面活性剂喷剂来维持生命，直到他们的肺发育成熟，能够产生足够的表面活性剂为止。

呼吸（即通气）是气体进出肺部的一种力学过程。正常成人每分钟大约呼吸 15～20次。吸气时，呼吸肌收缩，肺所处的胸腔容积扩大，从而导致肺泡增大，气体进入肺泡。随着肺泡的膨胀，呼吸系统内的压力降到比大气压约低 3mmHg，根据

Boyle 定律，此时气体就很容易进入肺部。呼气时，呼吸肌放松，胸腔恢复到原来的容积。由于呼吸系统的体积减小，其压力增大，约达到高于大气压 3mmHg，于是气体从肺部被排出，进入大气。

肺的工作原理与肺和胸壁的力学特性有关，肺的静力学是指肺容积保持恒定时肺的力学特性。要了解肺的力学特性首先需要知道肺的各种气量和容量，如图 3.25 所示，它们的含义如下：

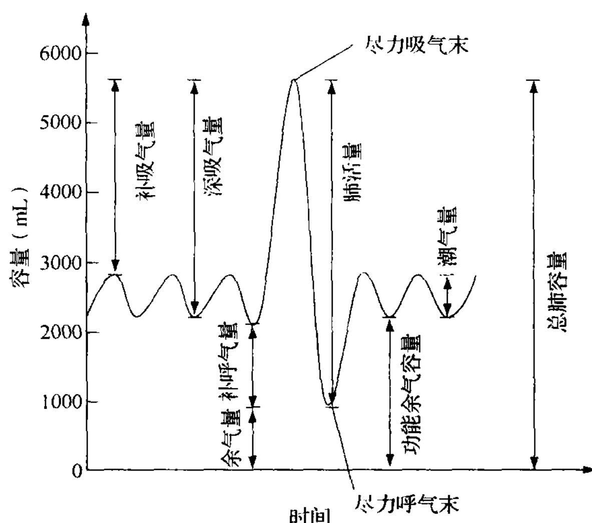


图 3.25 各种肺气量和肺容量

注：除了余气量、功能余气量和肺总容量之外，其他气量和容量都可以用肺量计测量。

潮气量（Tidal Volume，TV）——平静呼吸时肺部每次吸入或呼出的气量。

余气量（Residual Volume，RV）——尽力呼气末，肺中残存的气量。

补吸气量（Inspiratory Reserve Volume，IRV）——平静吸气末，再尽力吸气所能吸入的气量。

补呼气量（Expiratory Reserve Volume，ERV）——平静呼气末，再尽力呼气所能呼出的气量。

总肺容量（Total Lung Capacity，TLC）——尽力吸气末肺内的总气量。

肺活量（Vital Capacity，VC）——尽力吸气后，再尽力呼气所能呼出的最大气量。

吸气容量（Inspiratory Capacity，IC）——平静呼气末，再尽力吸气所能吸入的气量。

功能余气容量（Functional Residual Capacity，FRC）——平静呼气末肺内残余的气量。

除了几个包含余气量的指标以外，其他肺的各个气量和容量都可以用肺量计（spirometer）测得。常用的肺量计是一个充气容器，由一大一小两个圆桶组成。大

圆桶装水，底部有一根通气管通到水面上；小圆桶中有空气，倒过来覆盖在大桶内的通气管出口上，并浮在大桶的水中。通气管的另一端装有含口器，供病人使用。病人吸气时，浮桶内空气减少，浮桶下降；病人呼气时，浮桶就上升。记录浮桶位置的变化量，就可以计算出病人的各种肺气量。

例 3.9 假设某病人的肺总容量为 5.9L，如果用肺量计测得的吸气容量为 3.3L，那么该病人的功能余气容量应该是多少？如果要计算余气量，还需要测量什么指标？

解：由图 3.25 可知，肺总容量（TLC）等于吸气容量（IC）与功能余气容量（FRC）之和，即

$$TLC = IC + FRC$$
$$5.9L = 3.3L + FRC$$
$$FRC = 2.6L$$

由于

$$TLC - VC = RV$$

因此，要计算余气量（RV），就必须测出肺活量（VC）。 ■

肺量图（spirogram）可以记录肺容量随时间的变化过程，由此可以根据需要计算各种呼吸流速。例如，人尽力吸气到达肺总容量 TCL 之后，再尽力呼气到只剩余气量 RV。如果用 $FEV_{1.0}$（Forced Expiratory Volume）来表示开始尽力呼气 1 秒时所呼出的气量，那么，正常 $FEV_{1.0}$ 约为肺活量的 80%。患有限制性通气障碍的病人，其肺或胸壁的顺应性减小，或者呼吸肌衰弱，使吸气量减少，导致 $FEV_{1.0}$ 和肺活量都减小，但两者比率保持不变。患有哮喘等阻塞性通气障碍的病人，其 $FEV_{1.0}$ 下降幅度比肺活量的降幅大得多，这种病人的 TLC 异常大，但是呼气终止过早。还有一个有用的计算指标是尽力呼气流速 $FEF_{25\%\sim75\%}$（Forced Expiratory Flow），也就是在呼气中段（即达到肺活量 25%～75% 阶段）的平均流速。另外，流量—容积环（flow volume loop）是另一种分析肺功能的方法，它描述了吸气和呼气过程中气体的流速与流过的气体容积之间的关系。

肺总容量 TLC 可以用气体稀释法来测定。测量时，病人尽力吸入含有氦气等惰性示踪气体的混合气体（其中的氦气含量已知），达到 TLC 后，屏气 10s，使惰性气体在整个呼吸道和肺中分布均匀。根据质量守恒定律，吸入气体中示踪剂的浓度（已知）乘以吸气量（测得）应该等于呼出气体中示踪剂的浓度（从呼气中测得）乘以 TLC。注意，惰性气体不会通过肺进入循环系统。

体积描记法（plethysmography）是测量各种肺气量的最准确的方法。病人坐在空气密闭箱内，通过含口器呼吸。这种方法应用了 Boyle 定律，也就是在等温条件下，箱内气体压力和容积的乘积恒定不变。当病人呼吸时，根据肺容积的变化量和口腔压力的变化，就可以计算出功能余气容量。由于补呼气量可以测得，因此功能余气容量减去补呼气量就可以得到余气量。

例 3.10 假设某病人从一个含有 2L 混合气体的储气罐中呼吸，罐中含有 10% 惰性气体。呼吸足够长时间之后，储气罐和肺中的惰性气体达到平衡，此时测得罐

内惰性气体浓度变为 2.7%。那么，该病人的肺总容量是多少？

解：测量开始和结束时惰性气体的总量保持不变，但储气罐中的浓度由 10%（C_1）降到了 2.7%（C_2）。起始时，惰性气体只限于 2L（V_1）的储气罐内；结束时，储气罐和病人肺内都含有惰性气体，即 $V_2=V_1+TLC$。由此可得以下方程：

$$C_1V_1=C_2V_2$$

即

$$(0.1)(2L)=(0.027)(2L+TLC)$$

$$0.2L-0.054L=0.027TLC$$

$$TLC=5.4L$$

■

如图 3.26 所示，血液与肺泡之间进行的气体交换过程称为外呼吸（external respiration）。每个成人的肺大约含有 3.5×10^8 个肺泡，形成很大的气体交换表面积（约 60～70m^2）。每个肺泡只有一层细胞的厚度，使得气－血屏障总共只有两层细胞的厚度，即一层肺泡细胞加上一层毛细血管内皮细胞，约 2μm。肺泡中的氧分压高于血液中的氧分压，因此氧气从肺泡向血液扩散。而肺泡中的二氧化碳分压低于血液中的二氧化碳分压，因此二氧化碳从血液向肺泡扩散。血液与组织之间的气体交换过程称为内呼吸。在内呼吸过程中，二氧化碳和氧气在血液与体细胞周围的细胞外液之间扩散。气体扩散的方向和速度取决于血液和细胞外液中的气体分压大小、扩散表面积大小、气体需通过的膜的厚度以及扩散常数，其中扩散常数与气体溶解度和气体分子质量相关（即 Fick 定律）。

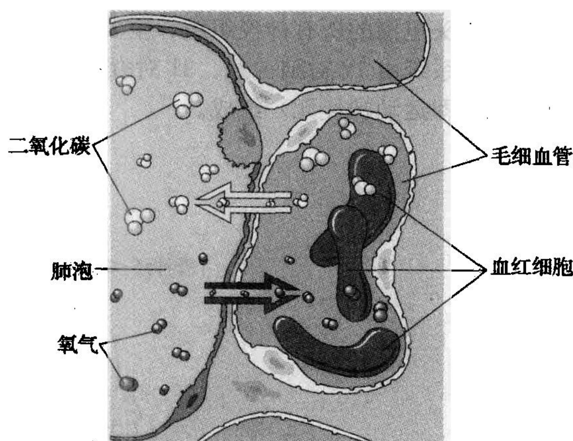


图 3.26 外呼吸过程中，氧气从肺泡进入血液，二氧化碳则从血液进入肺泡

机械呼吸机用于给病人提供空气或氧气，它们通常由微处理器控制，可以是电动产生气流，也可以用压缩空气供气。呼吸机可分为负压和正压两种类型。像铁肺之类的负压呼吸机是在胸腔周围产生负压，迫使气体进入肺部。这类负压呼吸机并不是任何病人都能使用。正压呼吸机则是向病人肺的入口处施加高压气体，使得气体或氧气沿着压力梯度进入肺部。呼吸机一般有两种通气模式，在“控制模式”下，呼吸机自动给病人产生呼吸周期；在“辅助模式”下，呼吸周期则由病人的自主呼吸启动，机器只是辅助通气。这种正压呼吸机在吸气期间会使胸腔内压升高，从而会影响静脉回流和心输出量。还有一种高频喷气式呼吸机以 60～900 次/min 的极高频率，将小容量气体脉冲送入肺部。此时，氧气和二氧化碳通过分子扩散原理进行交换，而不是通过大的气流进行交换。与正压呼吸机相比，这种高频呼吸机对心输出量的影响较小。另外，体外膜肺氧合（Extra Corporeal Membrane Oxygenation,

ECMO）利用体外循环技术，将病人的血液引入人工肺，完成氧气和二氧化碳的气体交换，然后，将血液加热到体温，再送回病人体内。这种技术使病人的肺得到休息并自我康复，在一些溺水者和患有可逆性肺病的婴儿身上已得到成功的应用。

3.4.3 神经系统

神经系统负责整合和控制全身各种功能，它分成两个主要部分：中枢神经系统（Central Nervous System，CNS）和外周神经系统（Peripheral Nervous System，PNS）（见图 3.27）。前者包括骨头包围着的所有神经组织，也就是脑和脊髓；后者包括无骨头包围的所有神经组织，主要用于感知和响应来自体内和体外的刺激信号。外周神经系统由 12 对脑神经、31 对脊神经以及传入神经元（即感觉神经元）和传出神经元（即运动神经元）组成。

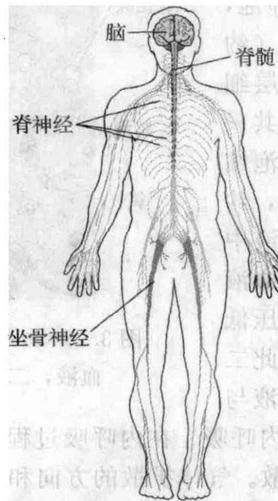


图 3.27　中枢神经系统由骨头包围着的所有神经组织组成，也就是脑和脊髓，而外周神经系统则由无骨头包围的所有神经组织组成

神经系统还可以分成躯体神经系统和自主神经系统，两个系统中都含有来自中枢神经系统和外周神经系统的部分。例如，躯体外周神经系统由感觉神经元和运动神经元组成，感觉神经元将疼痛、温度以及皮肤、肌肉和关节上的力学刺激等接受器的信息传给中枢神经系统；而运动神经元则将中枢神经系统的神经脉冲送回到身体的同一个区域。自主神经系统的功能主要是下意识调节平滑肌、心肌和各种腺体，它包括交感神经和副交感神经两部分。交感神经可以引起内脏和皮肤血管的收缩，骨骼肌血管的扩张，以及心率的增加；而副交感神经对内脏和皮肤的血管具有相反的作用，对骨骼肌无任何支配作用，并会引起心率减缓。因此，交感神经使人体处于一种“应急”状态，为“作战”或者“逃避”作准备，而副交感神经则使人体恢复到正常状态。

神经系统的各个部分都由传导电脉冲的神经元以及胶质细胞构成，胶质细胞具有保护、支持和营养神经元的作用。神经元胞体（见图 3.16d）的突起构成了一个轴突和多个树突分支，它们都由胞体供给营养。虽然胞体也能接受来自其他神经元的输入，但树突是神经元的主要接受器。通常每个神经元的树突要接受来自成千上万个与其他神经元形成的突触连接的信号。轴突将神经信号传递给大脑或者脊髓中的其他神经元，或者传递给外周腺体和肌肉。轴突长度的变化范围很大，从几毫米（大脑内）到近 1m（从脊髓到脚部）。有些轴突外面包裹着髓鞘，髓鞘由特异性细胞——许旺氏细胞（Schwann cell）构成。每个轴突的末端都有很多分支，称为突触前末梢。这些末梢的球形突起中含有许多容纳神经递质的突触囊泡。当神经元树突接受到刺激信号时，细胞膜对钠离子的通透性增加，就像心肌细胞中所发生的那样，神经元就会产生动作电位并沿着轴突传导出去。有髓鞘轴突上的髓鞘间隙（即郎飞氏结）可以使动作电位以更快的速度传播，也就是在郎飞氏结之间跳跃式传播。突触前末梢中的突触囊泡可以将所含的神经递质释放到轴突与邻近神经元、肌细胞或腺体细胞之间形成的突触间隙中，然后神经递质扩散到突触后膜上并引发响应（见图 3.28）。

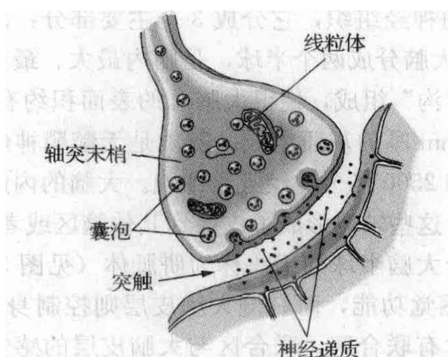


图 3.28　受到刺激之后，轴突末梢中的囊泡向突触移动，并通过胞吐方式将神经递质释放到轴突与突触后膜之间的间隙中，这个后膜可以是另一个神经元的树突、或者是肌纤维、或者是腺体细胞；神经递质通过扩散进入突触后膜并引发相应的细胞产生响应

神经元之间的相互连接有多种不同的类型。辐散式连接中，突触前神经元的轴突的每个分支都与不同的突触后神经元的树突形成突触连接。聚合式连接中，一个突触后神经元的树突上接受来自多个不同突触前神经元的轴突末梢并形成突触连接。直接反馈连接中，某个神经元的轴突与一个中间神经元的树突相连，该中间神经元的轴突又反过来与原神经元的树突相连。双神经元回路是指感觉神经元直接与运动神经元形成突触连接；而三神经元回路是指感觉神经元、脊髓中的中间神经元和运动神经元组成的连接。如图 3.29 所示，这两种回路在反射弧（reflex arc）中都存在。反射弧是一种特殊的神经回路，它的起始端是感受器（如指尖的痛觉感受器等）的感觉神经元，终末端是效应器（如骨骼肌等）的运动神经元。撤回反射主要由足以

引起疼痛感觉的痛刺激或热刺激引发，也称为保护性反射或回避反射。这种反射使人体能够对危险状况作出快速反应，省去了将信号传递给大脑分析处理后再传回的时间。

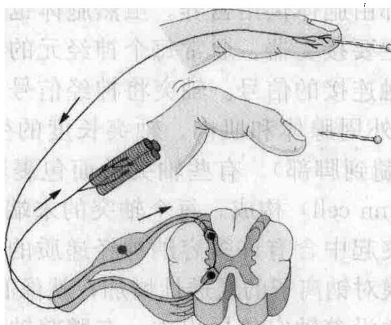


图 3.29　一个反射弧的例子

注：该反射弧的起始端是手指的感受器神经元，当指尖被针扎到时，感觉神经元产生动作电位，并传递给中间神经元，再传给运动神经元。运动神经元与手指肌纤维之间有突触连接，当兴奋刺激到达肌纤维时，肌纤维就会收缩，使指尖脱离针尖。

脑是一团大而柔软的神经组织，它分成 3 个主要部分：大脑、间脑以及脑干和小脑。如图 3.30 所示，大脑分成两个半球，是脑内最大、最显著的部分，它由许多隆起的“回”和狭窄的“沟”组成，使得大脑总的表面积约有 2.25m²。大脑的外层称为大脑皮层，由 2～4mm 厚的灰质组成，灰质是无髓鞘神经元。大脑皮层中约含有超过 500 亿个神经元和 2500 亿个神经胶质细胞。大脑的内层是较厚的白质，由一群群有髓鞘的轴突组成，这些轴突形成了皮层到其他脑区或者丘脑（间脑的一部分）到皮层的投射。连接两个大脑半球的组织称为胼胝体（见图 3.30b）。左侧大脑皮层控制身体右侧的运动和感觉功能，而右侧大脑皮层则控制身体左侧的运动和感觉。大脑左右两半球各自还含有联合区，联合区与大脑皮层的感觉区和运动区相连，它解释传入信号或者协调运动响应。

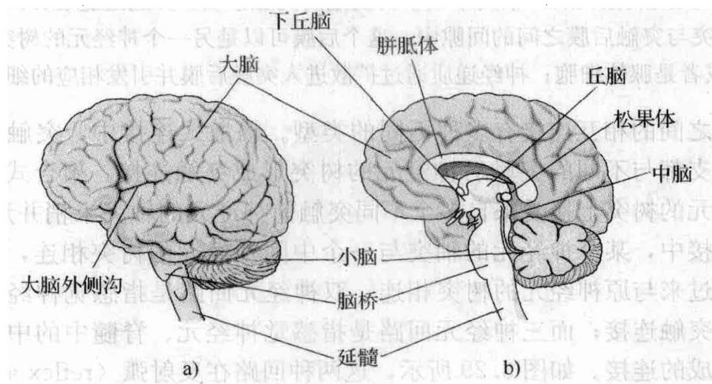


图　3.30

a）脑的外表面　b）脑的中矢状切面

大脑皮层的沟裂将每个大脑半球分成了许多具有不同功能的脑叶。其中，额叶（frontal lobes）的功能包括启动骨骼肌的随意运动、分析感受器的信息、产生与个性相关的反应，它还参与调节与记忆、情绪、推理、判断、计划和说话有关的反应。顶叶（parietal lobe）响应全身皮肤和肌肉感受器传上来的刺激信号。颞叶（temporal lobe）分析某些感受器的信息，存储听觉和视觉信息，它还包含了听觉中心，接受来自耳蜗感受器神经元的信息。枕叶（occipital lobe）通过控制眼睛的定向和聚焦来调整眼球的运动，并结合以往的视觉经验和其他感受器信息，修正视觉图像。脑岛（insula）是位于顶叶、额叶和颞叶下面的大脑深部的一部分，其功能目前还不太清楚，似乎与胃肠等内脏活动有关。

间脑是脑的深部组织，它将脑干的中脑区域与大脑半球相连接，主要包括丘脑、下丘脑和上丘脑（见图 3.30b）。丘脑与感觉系统、运动系统、常规神经背景活动、情绪的表达和独特的人类行为有关。由于它与大脑皮层区域之间具有双向联系，因此，它与思维、创造力、解释、口语和书面语的理解以及通过触觉识别物体等都有关。下丘脑与以下活动有关：自主神经系统信息的整合、体温的调节、水和电解质的平衡、睡眠一苏醒的模式、食物摄取、与情绪相关的行为反应、内分泌控制和性反应等。上丘脑含有松果体，后者被认为具有神经内分泌功能。

脑干连接脑和脊髓，自动控制呼吸等至关重要的人体功能。它主要包括中脑、脑桥和延髓（见图 3.30b）。中脑将脑桥、小脑与大脑相连，位于脑干的上端，它与视觉反射、眼球运动、晶状体聚焦以及瞳孔放大等有关。脑桥是位于中脑和延髓之间的圆形隆起，它与延髓一起控制呼吸功能，是延髓通向大脑高层结构的中继站，也是第五对脑神经的起始位置。延髓是脑干的最下端，它将脑桥与脊髓相连。延髓包含了许多重要的控制中心，调节心率、呼吸频率、血管收缩和扩张、血压、吞咽、呕吐、打喷嚏和咳嗽等。小脑位于脑桥背侧，是脑的第二大部分。它处理用于运动系统的感受器信息，并参与调节骨骼肌收缩和源于大脑皮层的随意肌肉运动信息。因此，小脑是调节平衡、维持身体姿势、保证身体运动的时空准确性的信息处理中心。

3.4.4 骨骼系统

成人一般共有 206 块骨头，但是骨头的实际数目人与人之间有所不同，并且随着年龄的增长，某些骨头会融合，使骨头的总数减少。就像人体分成主干和四肢两大部分一样，骨骼系统也分为主轴骨和四肢骨两大部分（见图 3.31）。主轴骨共有 80 块骨头，包括颅骨、舌骨、脊柱和胸廓等，四肢骨共有 126 块骨头，包括上肢带骨、下肢带骨、上肢骨和下肢骨等。骨骼系统起保护和支持身体的作用，并协助运动、制造血细胞、存储重要的矿物质。它由特殊结缔组织形成的骨头构成，骨头很强硬，承受着人体重量，是人体的主要支撑部件。某些软骨也具有一定的支撑作用，软骨是一种光滑、坚固、有弹性、无血管的结缔组织。骨骼上的骨头很坚硬，因此它们对骨骼包围之中的器官（如大脑和各种腹腔器官等）具有保护作用。

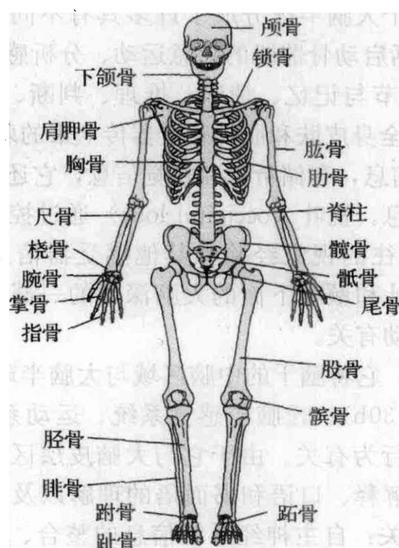


图 3.31 骨骼系统

注：舌骨（图中未显示）、脊柱和胸廓（包括肋骨、肋软骨和胸骨）构成了主轴骨；上肢带骨（包括肩胛骨和锁骨）、下肢带骨、上肢骨和下肢骨构成了四肢骨。

颅骨共有 29 块。其中大脑后部有 8 块颅骨包围、支撑并保护脑；还有 14 块面颅骨构成面部支架，给面部肌肉提供附着地，主要负责面部皮肤的活动。除下颌骨以外，其他面颅骨相互之间都紧密相连，并与脑颅骨相连。另外，听小骨共有 6 块，每只耳朵各 3 块，它们负责将外界环境的声波传入内耳。还有一块 U 形的舌骨，它靠近颅骨，但并不是颅骨的一部分，位于下颌之下的颈部，通过肌肉和韧带连接到颅骨和喉部，为重要的颈部和舌部肌肉提供附着。

幼儿的脊柱有 34 块骨，但成人平均只有 26 块独立的脊椎骨。其中，有 7 块颈椎骨，包括作为头部旋转枢纽的枢椎（即第 2 颈椎）、位于枢椎之上的寰椎（即第 1 颈椎，它支持整个头颅）以及紧随其后的 5 块颈椎骨；还有 12 块胸椎骨、5 块腰椎骨；最后是骶椎骨和尾骨各 1 块。骶椎骨和尾骨其实分别由幼年时的 5 块椎骨融合而成。脊柱支持头部和躯干的体重及其运动，保护脊髓，并构成椎间孔，使脊髓中分出的脊神经由此穿出。成人脊柱有 4 个弯曲，即颈曲、胸曲、腰曲和骶曲（也称尾曲），这种结构有利于躯体的弯曲，也可以减少震荡。虽然邻近两节椎骨之间的相对运动非常有限，但脊柱总体的动作范围却很大。胸廓由 12 块胸椎、12 对肋骨及其相应的肋软骨、胸骨组成，其中胸椎也是脊柱的组成部分。胸廓保护至关重要的器官，并防止呼吸时产生胸腔塌陷。

骨根据其形态可分为长骨、短骨、扁骨和不规则骨。股骨和肱骨等是长骨，其长度大于宽度。脚踝和手腕中的某些骨头是短骨，其长宽相近。胸骨和颅骨等是扁骨，具有相对较薄的平板结构。脊椎骨和髋骨等属于不规则骨，与其他类型骨头的

形状都不一样。

骨头重量约占全身体重的 18%，其密度约为 1.9g/cm³。骨组织有两种不同的类型：松质骨和密质骨。松质骨疏松多孔，构成长骨的末端（即骨骺）和其他骨头的内部结构。密质骨构成骨干和骨头的表层，其抗拉强度为 120N/mm²，抗压强度为 170N/mm²，杨氏模数为 1.8×104N/mm²。骨干中间的空腔（即骨髓腔）内充满脂肪状的黄骨髓或者红骨髓，后者含有造血细胞。

骨是一种一直在更新的活器官，旧的骨组织被一种特殊细胞（即破骨细胞）去除，新的骨头则由成骨细胞沉积形成。骨头的生长过程既可以使骨得到重新塑造，也可以调节人体对钙的需求。人的一生中全身骨骼平均要更新 3 遍。当旧骨降解速度比新骨生成的速度快时，就会导致骨头脆弱，从而容易发生骨折，这就是骨质疏松疾病。

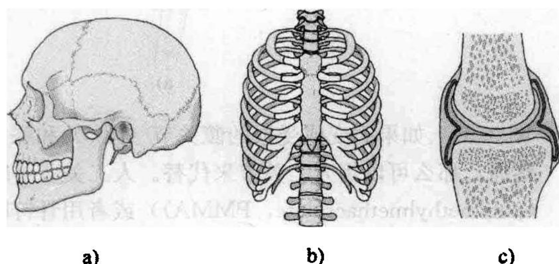


图 3.32 骨骼系统骨头之间的连接方式
a）纤维连结 b）软骨连结 c）关节连结

骨骼系统中各个骨头相互之间的连接方式有：纤维连结、软骨连结和关节连结（见图 3.32）。纤维连结通过纤维结缔组织将骨紧密地束缚在一起，这种连结刚性较大，骨头之间不能相对活动，或者只有很小幅度的活动，颅骨之间的缝连接就属于这种连结。软骨连结通过软骨将骨头连接起来。这种连接只能允许扭转和压缩等有限的活动，脊柱系统中肋骨与脊柱和胸骨之间的连接就属于这种连结。关节连结是最复杂多变的连接形式，关节腔内含有滑液，关节骨表面覆盖有软骨，还有韧带将各个部分连起来。膝盖就是一个关节连结。



图 3.33 关节中含有充液关节腔，是最复杂多变的一类骨连结，每个关节连结根据其结构和活动形式的不同可分成图中所示的 6 种类型

关节连结根据其结构和允许的运动形式分成 6 种类型。如图 3.33 所示，滑动关节是关节连结中最简单的一种，它可以进行前后或左右运动，如手腕的腕关节。肘等铰链关节只允许在一个平面上弯曲，是关节连结中最普遍的一种。寰椎与枢椎之间是轴承式的关节连结，允许旋转活动。踝状关节中，一块骨头的椭圆形凸起表面刚好嵌合在另一块骨头的凹陷处。掌骨关节就属于这种关节连结，它允许弯曲一伸展和旋转，属于一种双轴运动，能绕两个轴转动。拇指基部的连结属于鞍状关节，它是踝状关节的一种变形，可以绕多个轴转动。球一袋状关节允许绕某

个固定中心进行多方向的转动，这种关节中，一块骨头的球形头部正好与另一块骨头的杯形凹陷相互吻合。这种多轴关节是最灵活的关节，肩关节和髋关节就属于这种关节。如图 3.34 所示，生物医学工程人员已经开发了各种各样的人工关节，成为病变和受损的髋关节、肩关节和膝关节等关节的常规临床替代品。

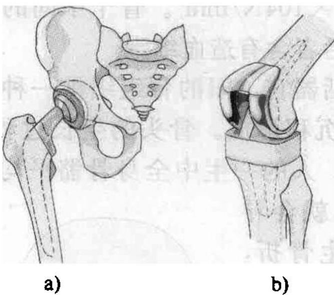


图 3.34 如果病变或受损的髋关节（图 a）和膝关节（图 b）丧失了功能，或者造成剧烈疼痛，那么可以用人工关节来代替。人工关节可以用特殊的胶合剂（即聚甲基丙烯酸甲酯（polymethylmethacrylate，PMMA））或者用骨钉固定；由于材料的弹性模量不同，如钛为 110GPa、PMMA 为 2.2GPa、骨头为 20GPa，因此连接界面上会产生一些特殊问题

3.4.5 肌肉系统

如图 3.35 所示，肌肉系统由 600～700 块骨骼肌组成，某些肌肉有时分别计数，有时成对计数，因此数目会有所不同。整个肌肉系统的重量占人体总重量的 40%左右。附着在主轴骨（见图 3.31）上的肌肉大约占人体总骨骼肌的 60%，它们保持头部和脊柱的姿势，并在呼吸时运动肋骨。四肢骨上附着的肌肉则运动固定四肢骨的各个组件。

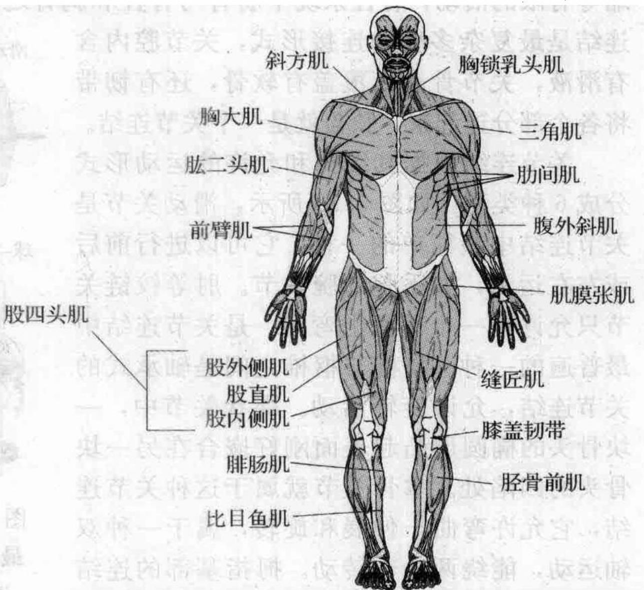


图 3.35 人体前面部分主要骨骼肌示意图

肌肉系统骨骼肌的功能是维持姿势、产热维持体温，并为骨头、关节和脸部皮肤的活动提供驱动力。在完成动作中起主要作用的肌肉称为主动肌，产生与主动肌相反作用的肌肉称为拮抗肌，协助主动肌的肌肉称为协同肌。某些骨骼肌的持续收缩可以维持身体的姿势，如果人发生晕厥，这些

肌肉松弛，那么，人就瘫倒在地。

如图 3.36 所示，骨骼肌的运动依据杠杆原理，即由绕支点转动的杠杆臂组成，每个杠杆上有两个作用力：一是需要移动的重量，也就是负载；二是施加的作用力。骨头就像杠杆臂，关节就是支点，需要克服的负载就是要移动的人体的某部分重量，而作用力由附着在骨头上的肌肉收缩产生，作用力的作用点就是肌肉在骨头上的附着点。杠杆系统可以分成 3 种不同的类型。第一类杠杆系统的支点位于作用力与负载之间，例如仰头时头部的运动就是这类杠杆，此时寰椎与颅骨枕骨之间形成支点，头后部的脊柱肌肉产生作用力抬起脸部的重量。第二类杠杆系统的负载位于作用力与支点之间。当人踮起脚尖站立时就是这种情况，此时前脚掌为支点，作用力由腿后部的小腿肌肉产生，需要移动的负载就是整个人体的重量。第三类杠杆系统的作用力位于负载与支点之间。当某人弯肘在身体前举起一个重球时就是这种情况，此时肱二头肌收缩产生作用力，需要移动的负载包括球的重量以及前臂和手的重量，肘关节是支点。

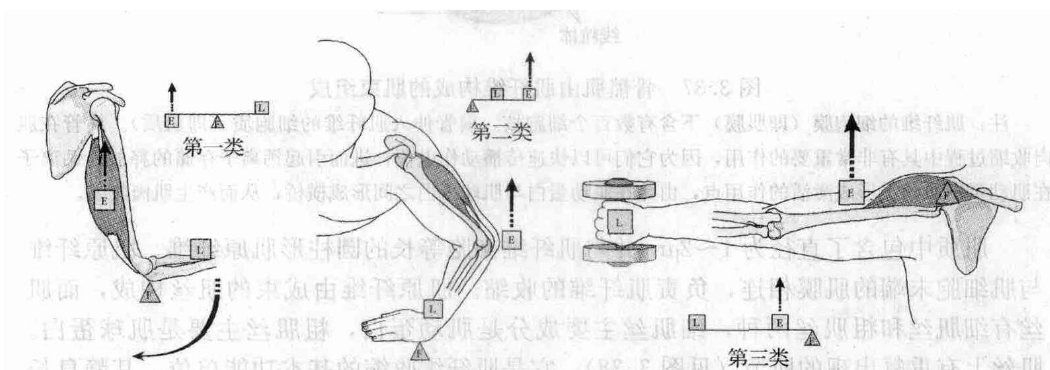


图 3.36　根据肌肉作用力的位置、负载的位置和支点的位置的不同，人体手臂可以构成 3 类不同的杠杆系统

注：图中 E 表示作用力，L 表示负载。

心肌、平滑肌、骨骼肌是 3 种不同的肌肉组织，但它们都具有 4 个重要的特性：收缩性，肌肉缩短的能力；兴奋性，肌肉接受和响应刺激的能力；展长性，肌肉拉伸的能力；弹性，肌肉拉伸或收缩之后恢复原状的能力。心肌只存在于心脏；平滑肌几乎在除了心脏之外的所有其他器官中都有，它们形成片状、束状或鞘状，包裹在其他组织周围；骨骼肌由骨骼肌的肌肉组织、结缔组织、血管和神经组织构成。

每一块骨骼肌外都有一层结缔组织（即胶原纤维）包裹着，它将肌肉与周围的其他组织和器官分开。这种胶原纤维在肌肉末端聚集在一起形成肌腱，骨骼肌就是通过肌腱与骨头、或者脸部皮肤、或者其他肌肉的肌腱（就像手的情况）连接起来。还有一种结缔组织纤维将骨骼肌分成一个个称为肌束的小部分，肌束中含有成束的肌纤维。每个肌束内，有别的结缔组织包裹每条骨骼肌纤维，并将相邻的肌纤维连在一起。每条骨骼肌纤维的细胞膜下包含有数百个细胞核。众多细胞核提供了多份基因备份，能

够指导生产肌肉正常收缩所需的各种酶和结构蛋白，由此可以加快收缩速度。

肌纤维的细胞膜和细胞质分别称为肌膜和肌质。如图 3.37 所示，横管（即 T 管）是由肌膜伸入肌质形成的，它垂直于肌膜表面。T 管具有协调收缩的作用，其内部是细胞外液，它在肌纤维中形成许多通道。T 管与肌质网（一种特殊的内质网）末端膨大部分（即终池）紧密接触。终池中包含了肌肉收缩所需的高浓度钙离子。

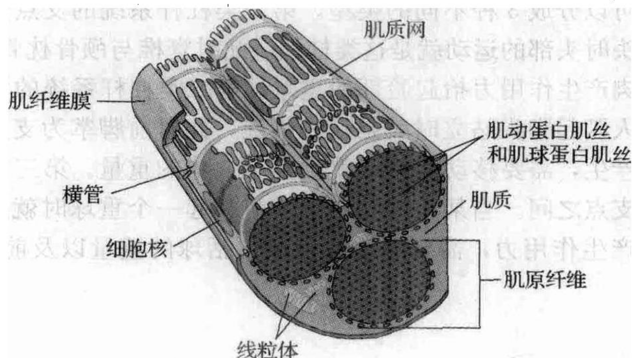


图 3.37 骨骼肌由肌纤维构成的肌束组成

注：肌纤维的细胞膜（即肌膜）下含有数百个细胞核。横管伸入肌纤维的细胞质（即肌质）。横管在肌肉收缩过程中具有非常重要的作用，因为它们可以快速传播动作电位，进而引起钙离子存储的释放。钙离子在肌动蛋白肌丝上形成激活的作用点，由此在肌动蛋白与肌球蛋白之间形成横桥，从而产生肌肉收缩。

肌质中包含了直径为 1～2μm 并与肌纤维细胞等长的圆柱形肌原纤维。肌原纤维与肌细胞末端的肌膜相连，负责肌纤维的收缩。肌原纤维由成束的肌丝构成，而肌丝有细肌丝和粗肌丝两种，细肌丝主要成分是肌动蛋白，粗肌丝主要是肌球蛋白。肌丝上有重复出现的肌节（见图 3.38），它是肌纤维收缩的基本功能单位，其静息长

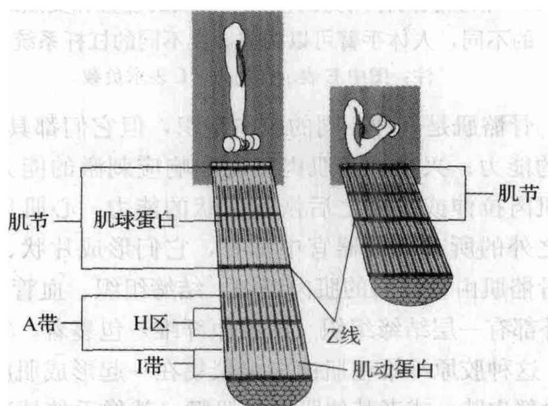


图 3.38 肌节是两条相邻 Z 线之间的区域，它是骨骼肌的基本功能单位。肌动蛋白肌丝附着于 Z 线，并伸向 A 带，在 A 带中它与较粗的肌球蛋白肌丝重叠。H 区是 A 带中不含肌动蛋白的一部分；当肌肉从伸展的松弛状态（左图）变成收缩状态（右图）时，肌球蛋白肌丝利用横桥拖动肌动蛋白肌丝，带动 Z 线之间相互靠近，从而使肌节缩短，肌肉收缩

度约为 2.6μm。其中，细肌丝附着于 Z 线，Z 线构成了肌节的两端。粗肌丝含有双头肌球蛋白，位于细肌丝之间。粗、细肌丝之间有部分重叠，因而使肌纤维呈现明暗交替的条纹。I 带是指放松状态下只包含肌动蛋白细肌丝的部分，H 区是指只含肌球蛋白粗肌丝的部分。H 区以及肌动蛋白与肌球蛋白重叠的部分一起构成了 A 带。

肌肉收缩时，粗肌丝的肌球蛋白分子与细肌丝肌动蛋白上的激活作用点之间形成横桥，将细肌丝拖向肌节中心；然后横桥解离，并在细肌丝的下一个作用点形成新的横桥连接。这就像用双手交替拉一根绳子一样。这种肌丝滑动机制消耗 ATP 能量，可以使肌肉缩短，肌肉的缩短（即肌肉收缩）就会带动其附带的骨头运动（见图 3.38）。

肌纤维与感觉神经元和运动神经元相连。感觉神经元末梢接受肌肉的张力和疼痛等信息，通过脊髓向大脑发送神经脉冲；而运动神经元末梢则接受来自大脑和脊髓的神经脉冲，并引起肌肉兴奋和收缩。每个运动神经元的轴突都有分支，可以支配多个肌纤维。其轴突分支没有髓鞘，并分成数个终端，与肌肉表面形成突触连接。当神经受到刺激时，轴突末端所含的囊泡就会将乙酰胆碱神经递质释放到神经元与肌肉之间的突触间隙中，乙酰胆碱扩散，并与肌膜运动终板上的受体结合，引起肌膜钠离子通道开放，肌纤维就产生动作电位。该动作电位沿着整个肌膜表面传播开来，并传到所有 T 管中，触发终池快速释放大量钙离子。钙离子诱发细肌丝上作用点的产生，并与肌球蛋白之间形成横桥，于是引起肌肉收缩。随着收缩过程的进行，乙酰胆碱酶逐渐分解乙酰胆碱，使得肌膜对钠离子的通透性恢复到原有的低水平，肌肉收缩也就结束了。

运动神经元及其支配的肌纤维组成一个运动单位。一个运动单位中的所有肌纤维都一起收缩，而同一块肌肉中属于不同运动单位的肌纤维的收缩时间可以不同。当收缩的肌肉松弛时，在其他收缩肌肉的牵动下或在重力作用下，肌肉会恢复到原始静息长度。肌肉放松期间，ATP 用于将钙离子泵回终池，横桥形成所需的作用点也被盖住，因此肌动蛋白和肌球蛋白脱离。横桥消失后，肌肉恢复到静息的松弛长度。

人体内有快、慢两类骨骼肌纤维。快肌纤维在接受刺激后 10ms 或更短的时间内就能收缩，人体内绝大多数骨骼肌都属于这种类型。快肌纤维半径大，含有密集的肌原纤维、大量用于生产 ATP 的糖原储备以及较少的线粒体。这种纤维能产生有力的收缩，消耗大量 ATP，因此会迅速产生疲劳，也就是即使受到不断的神经刺激也无法继续收缩。相反，慢肌纤维发生收缩响应的时间大约是快肌纤维的 3 倍，但它们能长时间持续收缩，因为它们含有比较丰富的毛细血管网络，可以摄取更多的氧气；含有一种特殊的氧结合蛋白分子，称为肌红蛋白；含有较多线粒体，产生的 ATP 比快肌纤维多。各种肌肉中快肌纤维和慢肌纤维的含量各不相同。主要成分是快肌纤维的肌肉（如鸡胸脯肉）看起来比较白，而主要成分是慢肌纤维的肌肉（如鸡腿肉）则比较红。人体肌肉一般是两种肌肉混合物，因此呈粉红色。每块肌肉所含快肌纤维和慢肌纤维的比例由基因决定，但通过运动训练可以提高快肌纤维的抗疲劳能力。

3.5 内环境的自身平衡

人体各种器官系统的协同工作维持着体内环境的稳定。内环境的自身平衡（homeostasis）是指外界环境发生变化时，人体内部环境中的物理和化学状态仍能维持在允许范围之内。例如，体温、血压、呼吸频率和心率等都是由自身平衡机制调控的，自身平衡是多个器官系统共同作用的结果。

细胞外液是细胞在体内所处的环境，对于体内自身平衡起着非常重要的作用。它在全身循环，将各种物质输送给细胞或者从细胞带走；它也是维持最佳体温和血压水平、维持体内酸碱平衡以及氧、二氧化碳、水、营养物质等许多化学物质浓度平衡的机制。

如图 3.39 所示，自身平衡由感受器、整合器和效应器 3 部分的相互作用来完成。感受器可以是细胞或者是细胞的某个部分，它检测环境变化所产生的刺激信号，并将刺激信息传送给整合器。整合器是控制中心，收集各个感受器传来的信息，然后诱发效应器产生响应。大脑就是一个整合器，它可以将整合所得信息发送给各个肌肉组织或者腺体，从而引起这些效应器产生特定的反应。大脑的整合功能就是将接收到的人体各个部分的实际运行信息，与人体各部分应该达到的运行状态进行比较，然后输出控制信息。

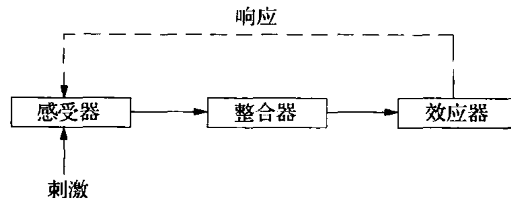


图 3.39 人体通过反馈机制维持自身平衡

注：感受器接收到刺激信号之后，或者直接将信息传送给效应器，或者将信息传送给整合器，再由整合器控制效应器产生响应。效应器的响应反馈到感受器，从而调节刺激对感受器的作用。对于负反馈，其反馈信号是削弱刺激对感受器的作用；而对于正反馈，其反馈信号则增强刺激对感受器的作用。

如果人体对于初始刺激的响应会进一步增强刺激效果，这种机制是正反馈机制，由于这种情况会破坏机体的自身平衡，因此在体内很少见。孕妇分娩过程就是一个正反馈过程，产道内婴儿头部对子宫颈感受器产生压力刺激，感受器将这种信号传送到下丘脑。下丘脑响应刺激，释放催产素，从而增强子宫收缩。子宫收缩强度的增加进一步推动婴儿在产道中前行，这更增强了子宫颈感受器的压力刺激。这种正反馈过程一直要持续到胎儿出生为止，这时，子宫颈感受器上的压力作用消失，下丘脑也就不再受到刺激，不释放催产素。

如果人体对于初始刺激的响应会减弱刺激的效果，引起相反方向的作用，这就是负反馈机制。例如，皮肤等人体各个部位的温度感受器检测体温，并将体温信息

传递到脑内的下丘脑部位，在那里与人体应该保持的温度（约37℃）进行比较。如果体温过低，下丘脑就发出信息使靠近皮肤的血管平滑肌收缩，从而减小血管直径，减少皮肤的散热；同时控制骨骼肌强烈收缩，即发生颤栗，立毛肌（见图3.16a）使汗毛竖起，形成一个个"鸡皮疙瘩"。这些肌肉收缩的代谢活动能够产生热量，从而温暖身体相反，如果体温过高，下丘脑就发送信息使邻近皮肤的血管平滑肌松弛，从而使血管直径增大，皮肤散热增加；同时控制汗腺分泌汗液，由此增加蒸发作用，冷却皮肤。当循环中的血液温度按照某个方向变化到人体内的设定值时，下丘脑就不再向肌肉和腺体等效应器发送信息。

再举一个人体负反馈机制的例子，就是胰腺的胰岛细胞对血糖水平的调节作用（见图3.40）。成人胰腺中大约分散着$2\times10^5\sim2\times10^6$个胰岛。当葡萄糖浓度升高时，胰岛的β细胞会产生胰岛素，从而促进葡萄糖的跨膜转运，使进入细胞的葡萄糖增加，并且加速葡萄糖转化成为糖原，存储在肝脏内。在饥饿等状态下，当血糖水平降到低于正常的70～110mg/dL水平时，胰岛的α细胞就产生胰高血糖素，刺激肝脏

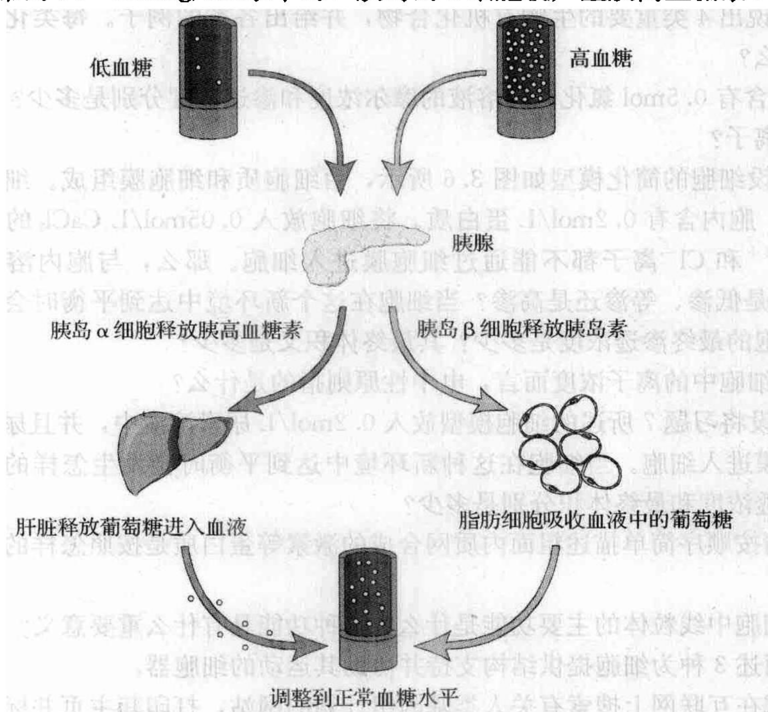


图3.40 控制血糖水平的两个负反馈机制

注：当血糖水平高于人体设定值时，产生刺激信号，刺激胰岛中的β细胞（即感受器）产生胰岛素（即信号），从而促进葡萄糖的跨膜转运并加速葡萄糖转化为糖原，存储在肝脏（即效应器）内，由此使得血液中的葡萄糖浓度下降。当血糖水平降至体内设定值时，β细胞就停止产生胰岛素。当血糖水平低于体内设定值时，又产生刺激，胰岛中的α细胞（即感受器）产生胰高血糖素（即信号），刺激肝脏（即效应器）将糖原转化成葡萄糖，因而使得血糖水平升高。直至血糖水平升至体内设定值时，α细胞就停止产生胰高血糖素。

将糖原转化为葡萄糖，并将其他氨基酸和乳酸等非碳水化合物转化成葡萄糖。当血糖水平恢复正常时，胰岛中的效应细胞就不再产生相应的激素（即胰岛素或胰高血糖素）。生物医学工程人员正在研究一种能够检测血糖水平，并模仿胰岛细胞的响应过程的药物输送系统，他们也在开发一种能有效地维持正常血糖水平的人工胰腺。

习　　题

1. 请运用解剖学术语，表述嘴巴与左耳、鼻子和右脚大脚趾之间的位置关系。

2. 请运用解剖学术语，表述胃在人体内的位置，以及它与心脏的相对位置关系。

3. 试在互联网上搜索一幅计算机断层扫描成像（CT）或者核磁共振（MRI）成像技术得到的人体横切面图，打印该图像并标出网址。

4. 试在互联网上搜索一幅 CT 或者 MRI 成像技术得到人体前视图，打印该图像并标出网址。

5. 请说出 4 类重要的生物有机化合物，并给出各类的例子。每类化合物的主要功能是什么？

6. 1L 含有 0.5mol 氯化钙的溶液的摩尔浓度和渗透浓度分别是多少？溶液中含有多少个氯离子？

7. 假设细胞的简化模型如图 3.6 所示，由细胞质和细胞膜组成。细胞的初始体积为 2nL，胞内含有 0.2mol/L 蛋白质。将细胞放入 0.05mol/L $CaCl_2$ 的大容积溶液中，且 Ca^{2+} 和 Cl^- 离子都不能通过细胞膜进入细胞。那么，与胞内溶液相比，该 $CaCl_2$ 溶液是低渗、等渗还是高渗？当细胞在这个新环境中达到平衡时会发生怎样的变化，细胞的最终渗透浓度是多少？其最终体积又是多少？

8. 就细胞中的离子浓度而言，电中性原则指的是什么？

9. 假设将习题 7 所述的细胞模型放入 0.2mol/L 尿素溶液中，并且尿素分子可以穿过细胞膜进入细胞。当细胞在这种新环境中达到平衡时会发生怎样的变化？细胞的最终渗透浓度和最终体积分别是多少？

10. 请按顺序简单描述粗面内质网合成的激素等蛋白质是按照怎样的徐径到达细胞外的。

11. 细胞中线粒体的主要功能是什么？这种功能具有什么重要意义？

12. 简述 3 种为细胞提供结构支持并协助其运动的细胞器。

13. 试在互联网上搜索有关人类基因组计划的网站，打印其主页并标出网址。在该网站查找并打印含有引起囊性纤维化的基因的染色体组型图。

14. 简述复制和转录的主要区别。

15. 简述细胞 DNA 基因所含遗传信息是如何表达成蛋白质，从而指导细胞活动的。

16. 亮氨酸有 6 个不同的密码子，而蛋氨酸只有一个密码子，这种编码方式对于调控蛋白质的翻译和生产具有什么重要意义？

17. 胰岛素是第一个被测序的蛋白质，其蛋白质序列如图 3.41 所示，假设其中没有内含子，那么产生分子中最后 4 个氨基酸的 DNA 序列可能是怎样的?

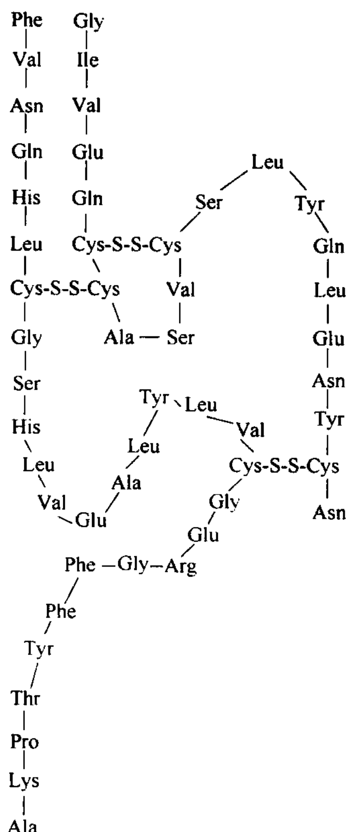


图 3.41　牛胰岛素由两条通过二硫键结合的肽链构成，两条链之间以及同一条链的不同部分之间还存在着氢键结合，其中不同的氨基酸用三字母名称表示（见表 3.1）

18. 请查找 5 篇经过同行评审的学术期刊的论文，内容是有关工程技术在 5 种不同人体器官系统中的应用，一篇论文涉及一个系统。请摘录论文的标题页信息及摘要。不能是综述、会议论文、主题发言讲稿等，也不能是书籍的章节、大众化出版物和报纸、社论等。较合适的查找期刊有《生物医学工程纪事》(Annals of Biomedical Engineering)、《IEEE 生物医学工程汇刊》(IEEE Transactions on Biomedical Engineering)、《IEEE 医学与生物工程杂志》(IEEE Engineering in Medicine and Biology Magazine) 以及《医学和生物工程学及计算》(Medical and Biological Engineering and Computing) 等。从论文中的什么信息可以看出论文是同行评审的?

19. 血红细胞从右手毛细血管网流动到右肺毛细血管网再返回，需要经过哪些途径? 在此过程中交换了哪些气体? 是在何处交换的?

20. 请画出肺循环和体循环的血流框图，其中包括心脏的各个房室、瓣膜、进出

心脏的主要动脉和静脉、肺以及体毛细血管网，并用箭头标出每个部分的血流方向。

21. 试在互联网上搜索一段正常窦性心律的 ECG，并用它说明心率的计算方法。

22. 为什么用 R 波（见图 3.22）而不是 T 波来计算心率？

23. 如果用热稀释法测定心输出量，如何计算？

24. 某高血压患者的收缩压为 145mmHg，舒张压为 98mmHg，那么其脉搏压和平均动脉压应该分别是多少？

25. 某病人的肺总容量为 5.5L，如果其肺活量为 4.2L，潮气量为 500mL，补呼气量为 1.2L，请计算补吸气量。

26. 如果要计算例 3.10 中病人的余气量，还需要测量哪些数据？

27. 请简述中枢神经系统、外周神经系统、躯体神经系统、自主神经系统、交感神经系统以及副交感神经系统的功能和主要组成部分，其中哪些神经系统隶属于其他神经系统？

28. 请解释肌节是如何缩短的，它如何引起肌肉收缩。

29. 肌肉系统和骨骼系统是如何相互作用产生运动的？

30. 请画出框图说明血糖水平调节的负反馈机制，并标出输入、感受器、整合器、效应器和输出。

推荐阅读资料

Brown, B.H., Smallwood, R.H., Barber, D.C., Lawford, P.V. and Hose, D.R. (1999). *Medical Physics and Biomedical Engineering*. Institute of Physics Publishing, Bristol and Philadelphia.

Cooper, G.M. (2000). *The Cell—A Molecular Approach*, 2nd Ed. ASM Press, Washington, D.C.

Deutsch, S. and Deutsch, A. (1993). *Understanding the Nervous System: An Engineering Perspective*. IEEE Press, New York.

Fox, S.I. (2004). *Human Physiology*, 8th Ed. McGraw-Hill, Boston.

Germann, W.J. and Stanfield, C.L. (2005). *Principles of Human Physiology*, 2nd Ed. Pearson Benjamin Cummings, San Francisco.

Guyton, A.C. (1991). *Basic Neuroscience: Anatomy & Physiology*. W. B. Saunders, Philadelphia.

Guyton, A.C. and Hall, J.E. (2000). *Textbook of Medical Physiology*, 10th Ed. W. B. Saunders, Philadelphia.

Harold, F.M. (2001). *The Way of the Cell: Molecules, Organisms and the Order of Life*. Oxford Univ. Press, New York.

Karp, G. (2002). *Cell and Molecular Biology—Concepts and Experiments*, 3rd Ed. Wiley, New York.

Katz, A.M. (1986). *Physiology of the Heart*. Raven Press, New York.

Keynes, R.D. and Aidley, D.J. (1991). *Nerve & Muscle*, 2nd Ed. Cambridge Univ. Press, Cambridge.

Leff, A.R. and Schumacker, P.T. (1993). *Respiratory Physiology: Basics and Applications*. Saunders, Philadelphia.

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D. and Darnell, J. (2000). *Molecular Cell Biology*, 4th Ed. W. H. Freeman, New York.

Martini, F.H. (2001). *Fundamentals of Anatomy & Physiology*, 5th Ed. Prentice Hall, Upper Saddle River, NJ.

Matthews, G.G. (1991). *Cellular Physiology of Nerve and Muscle*. Blackwell Scientific, Boston.

Pollack, G.H. (2001). *Cells, Gels and the Engines of Life—A New Unifying Approach to Cell Function*. Ebner & Sons, Seattle, WA.

Rhoades R. and Pflanzer, R. (2003). *Human Physiology*, 4th Ed. Thomson Learning, Pacific Grove, CA.

Silverthorn, D.U. (2004). *Human Physiology—An Integrated Approach*, 3rd Ed. Pearson Benjamin Cummings, San Francisco.

Tözeren A. and Byers, S.W. (2004). *New Biology for Engineers and Computer Scientists*. Pearson Education, Upper Saddle River, NJ.

Van De Graaff, K.M., Fox, S.I. and LaFleur, K.M. (1997). *Synopsis of Human Anatomy & Physiology*. Wm. C. Brown, Dubuque, IA.

West, J.B. (1990). *Respiratory Physiology—The Essentials*, 4th Ed. Williams & Wilkins, Baltimore.

Widmaier, E.P., Raff, H. and Strang, K.T. (2004). *Vander, Sherman, & Luciano's Human Physiology—The Mechanisms of Body Function*. McGraw-Hill, Boston.

第4章　生物力学

Joseph L. Palladino 博士
Roy B. Davis 博士

本章目录

本章主要内容：

- 运动学理论在生物力学中的应用；

• 动力学理论在生物力学中的应用；
• 材料力学在各种生物结构中的应用；
• 利用 MATLAB 程序求解生物力学中的静力学方程和动力学方程；
• 利用 Simulink 程序研究生物组织的粘弹性；
• 运动学方程在临床步态分析中的应用；
• 动力学方程在临床步态分析中的应用；
• 如何应用生物力学方法定量分析病人步态的病理状况，提出临床治疗方案，并定量分析治疗效果；
• 生物流体的流变学基本理论；
• 描述血管力学特性的模型；
• 心脏力学的基本理论；
• 生物力学在心血管系统中的应用，包括心脏泵血功能的定量分析、心脏与血管之间相互作用的研究以及各种临床应用的开发。

4.1 绪论

生物力学将经典力学原理应用于生物系统的研究，它是工程学与生命科学相结合所产生的一门新兴交叉学科。其研究范围很广，包括生物材料、心血管和呼吸系统的生物流体力学、材料的特性、医用植入物与机体之间的相互作用、生物组织（如肿瘤等）的热传输和物质传输、调节新陈代谢和随意运动的生物控制系统、人体步态分析的运动学和动力学应用等。各种生物机体及其系统的复杂性和特异性使生物力学领域涵盖的研究范围非常广泛。

本章有两个目标：一是介绍基本工程学原理在各种生物结构中的应用；二是介绍各种临床应用的开发。本章 4.2 节将复习有关静力学和动力学的基本原理；在此基础上，4.3 节将介绍材料力学的基本原理，这些都是工程人员应该掌握的基础知识；4.4 节将介绍生物材料的粘弹性；4.5 节则介绍生物力学原理在软骨、韧带、肌腱和肌肉中的应用。最后两节将综合应用这些基本知识，论述人体步态分析和心血管动力学研究这两大具体的临床研究方向。

人体就像一台非常复杂的机器，其骨骼和韧带系统构成了机架，肌肉和肌腱相当于电动机和电缆。人体的步行在力学上可以看做由杠杆（即骨）和轴承（即关节）组成的结构（即骨骼）在运动，其中的力由成对的主动肌和拮抗肌产生。因此，结构的强度和肌肉的作用是人体步行的重要基础。同理，在力学上，心血管系统可以看作一台复杂的泵（即心脏）将一种复杂的流体（即血液）泵入一组复杂的管道（即血管）。本书将只介绍步态分析和心血管动力学方面最基本的知识，读者如果需要深入了解有关的详细内容，请参阅本章末尾所列的参考文献。

力学具有悠久的发展历史，其最早的记载始于公元前 384－322 年古希腊亚里士

多德时代，那时，希腊人创立了正确的静力学理论，但是，描述生命系统的那些基本动力学理论却直到中世纪末新世纪初才开始建立。从 16 世纪开始，在 Kepler、Galileo、Descartes、Huygens 和 Newton 等人的努力下，动力学研究迅速发展。后来，Euler、LaGrange 和 LaPlace 等人将各种动力学定律进一步完善并系统化（参见 Dugas 所著的“力学发展史”）。

1638 年 Galileo 出版了《关于两门新科学的对话》（Dialogues concerning two new sciences）一书，书中介绍了力、运动和材料强度的概念。从此，“力学”就广泛用于描述所有系统的力和运动，包括量子、原子、分子、气体、液体、固体、结构、行星和星系等。生物系统也自然而然地成为力学的研究对象。

尚属新兴学科的生物力学就是应用力学原理研究生命系统的科学。著名的生物力学教授冯元桢（Y. C. Fung）博士这样描述生物力学在生物学、生理学和医学中的作用：

“如果没有生物力学，就不能深入了解生理学；就像没有空气动力学，就不能研制飞机一样。力学使人们能够设计飞机，并预测其性能；同样，生物力学有助于人们了解生理器官的正常功能，预测其变化，并设计人工干预的方法。因此，医学的诊断、手术和修复都离不开生物力学。”㊀

显然，生物力学是评估和改善人体健康的基础。

以下所列是在生物力学发展史上作出重要贡献的人物及其里程碑式的事件，尤其是那些与本章所述内容相关的发展历史：

Galen of Pergamon（129-199）发表了许多医学著作，包括《论肌肉的运动》（On the Movements of Muscles），他认识到运动需要肌肉的收缩。

Leonardo da Vinci（1452-1519）最早精确地描述了肩关节和髋关节等球窝关节，并将髋关节称为“人体的支柱”。他作图描绘了沿肌肉纤维方向的机械作用力。

Andreas Vesalius（1514-1564）撰写了《人体的结构》（The Fabric of the Human Body）一书。他以尸体解剖为基础，比 Galen 更精确地描述了人体肌肉组织的解剖结构，并证明运动来源于肌肉的收缩，且肌肉收缩时会变粗变短。

Galileo Galilei（1564-1642）致力于医学和物理学研究，将测量和观察相结合，并认为数学是科学的基础。他的研究内容包括：跳跃运动的生物力学、马和昆虫的步态以及动物骨骼的结构等。

Santorio Santorio（1561-1636）运用 Galileo 的测量和分析方法，发现了人的体重会随时产生变化，并由此开创了新陈代谢的研究。他是医学科学研究的先驱。

William Harvey（1578-1657）设计了一种实验方法，论证动脉与静脉之间存在闭合通路，为现代循环系统理论的建立打下了基础。1661 年，Malpighi 发现了血液循环系统的基本组成部分——毛细血管，从而证实了 Harvey 的推断。

Giovanni Borelli（1608-1679）是一位数学家，研究肌体动力学、肌肉收缩、动

㊀ 摘自 Biomechanics：Mechanical Properties of Living Tissues，Y. C. Fung，1993。——原文注

物运动以及心脏和肠的运动等，1680 年他发表了著作《论动物的运动》(On the Motion of Animals)。

Jan Swammerdam (1637-1680) 发明了神经—肌肉实验标本，并证实通过刺激连接在蛙腿肌肉上的神经，可以引起肌肉的收缩。他还证明，肌肉收缩时其体积会有所变化，驳斥了前人有关“动物灵魂”使肌肉收缩的说法。他的思想曾经引起轰动。

Robert Hooke (1635-1703) 创立了 Hooke 定律，阐明了弹性材料的应力与应变之间的关系，并在生物学中最先使用“细胞”这个名词。

Isaac Newton (1642-1727) 虽然在生物力学领域没有什么贡献，不过，他建立了微积分学、经典运动定律以及粘性流体的本构方程，这些都是生物力学的基础。

Nicholas André (1658-1742) 在他 80 岁时创立了整形术，他认为肌肉的失平衡是导致骨骼畸变的原因。

Stephen Hales (1677-1761) 可能是最早测量血压的人，在他 1733 年写的著作中包含了关于动物血液和血管压力实验的论述。

Leonard Euler (1707-1783) 将牛顿运动定律推广到连续介质，并用于描述刚体的运动，用于动脉中脉搏波的研究。

Thomas Young (1773-1829) 研究了振动与声音、光波理论与视觉，并提出了弹性力学理论的杨氏模量。

Ernst Weber (1795-1878) 和 Eduard Weber (1806-1871) 兄弟俩在 1836 年发表了《论人体步态力学》(On the Mechanics of the Human Gait Tools)，是最早研究人体步态的人。

Hermann von Helmholtz (1821-1894) 研究的问题非常广泛，包括光学、声学、热力学、电动力学、生理学和医学，还包括检眼镜、神经传导速度的测量、肌肉收缩的产热等。

Etienne Marey (1830-1904) 研究了马、鸟、昆虫、鱼和人的运动。他发明了测量地面反作用力的方法，还发明了电影摄影机。

Wilhelm Braune 和 Otto Fischer 两人在 1895～1904 年期间发表了著作《人体的步态》(The human gait)，对人体行走进行了数学分析，他们提出的方法至今仍在使用。他们还发明了轮转式全景摄影机，最早应用多部摄影机同时拍摄的方法来重建三维运动模型，并且应用牛顿力学的原理计算了关节的作用力以及肢体的加速度。

4.2　基础力学

本节将复习力学导论课中学过的某些要点，包括静力学和动力学。这方面的参考书有很多，如 Merriam 和 Kraige2002 年所著的《工程力学》(Engineering Mechanics) 就是一本很好的参考书。下面先复习矢量数学，然后是坐标变换矩阵（有些同学可能还没学过）、静力学平衡、转动惯量等，4.2.5 节将介绍的欧拉运动方程对于有些同学可能也是新内容。有关坐标变换和运动方程这两方面的内容可以参考

Greenwood的著作《动力学原理》(Principles of Dynamics)。

4.2.1 矢量数学

力可以用标量和单位矢量来表示，或者在极坐标中用幅值和方向来表示。如图4.1所示，二维矢量 $\boldsymbol{F}$ 可以分解为 x 方向的 $\boldsymbol{i}$ 分量 F_x 和 y 方向的 $\boldsymbol{j}$ 分量 F_y，即

$$\boldsymbol{F}=F_x\boldsymbol{i}+F_y\boldsymbol{j} \tag{4.1}$$

如 $20\boldsymbol{i}+40\boldsymbol{j}$lb⊖。本章中的矢量都用黑体字母表示。矢量也可以写成极坐标的形式，用矢量的幅值 $|\boldsymbol{F}|$（也称模）和矢量角 θ 来表示，即

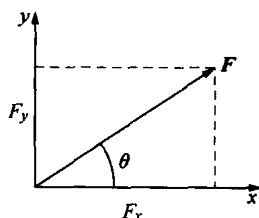


图4.1 矢量 $\boldsymbol{F}$ 的二维表示法

$$|\boldsymbol{F}|=\sqrt{F_x^2+F_y^2} \tag{4.2}$$

$$\theta=\arctan\frac{F_y}{F_x} \tag{4.3}$$

于是，矢量 $20\boldsymbol{i}+40\boldsymbol{F}$lb 的 $|\boldsymbol{F}|=44.7$lb，$\theta=63.4°$。同理，矢量可以用三维空间的 $\boldsymbol{i}$、$\boldsymbol{j}$ 和 $\boldsymbol{k}$ 分量来表示，即

$$\boldsymbol{F}=F_x\boldsymbol{i}+F_y\boldsymbol{j}+F_z\boldsymbol{k} \tag{4.4}$$

$\boldsymbol{k}$ 为 z 方向的分量。

通常，如果已知矢量的大小及其作用线经过的两个坐标点，例如，假设图4.2所示矢量 $\boldsymbol{F}$ 的大小为10lb，它经过原点（0，0，0）和点（2，6，4）；那么，$\boldsymbol{F}$ 就可以写为其大小 $|\boldsymbol{F}|$ 与单位矢量 $\boldsymbol{e}_F$ 的乘积，$\boldsymbol{e}_F$ 的方向与 $\boldsymbol{F}$ 的方向相同，即

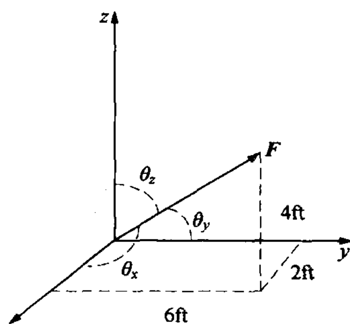


图4.2 由矢量的大小及其作用线定义的三维矢量

$$\boldsymbol{F}=|F|\boldsymbol{e}_F$$
$$=10\text{lb}\left(\frac{2\boldsymbol{i}+6\boldsymbol{j}+4\boldsymbol{k}}{\sqrt{2^2+6^2+4^2}}\right)$$
$$\boldsymbol{F}=2.67\boldsymbol{i}+8.02\boldsymbol{j}+5.34\boldsymbol{k}\text{lb}$$

圆括号中的量就是 $\boldsymbol{F}$ 的单位矢量 $\boldsymbol{e}_F$，即

$$\boldsymbol{e}_F=\left(\frac{2\boldsymbol{i}+6\boldsymbol{j}+4\boldsymbol{k}}{\sqrt{2^2+6^2+4^2}}\right)=0.267\boldsymbol{i}+0.802\boldsymbol{j}+0.534\boldsymbol{k}$$

且 $\boldsymbol{F}$ 的大小为

$$|\boldsymbol{F}|=\sqrt{2.67^2+8.02^2+5.34^2}=10\text{lb}$$

图4.2所示的矢量 $\boldsymbol{F}$ 也可以用三维空间中矢量作用线与各坐标轴之间的夹角来表示。假设矢量 $\boldsymbol{F}$ 与 x、y 和 z 轴之间的夹角（称为方向角）分别为 θ_x、θ_y 和 θ_z，则

$$\cos\theta_x=\frac{F_x}{|\boldsymbol{F}|} \tag{4.5}$$

⊖ "lb"是单位"磅"。——译者注

$$\cos\theta_y = \frac{F_y}{|\boldsymbol{F}|} \tag{4.6}$$

$$\cos\theta_z = \frac{F_z}{|\boldsymbol{F}|} \tag{4.7}$$

这些比值称为矢量 $\boldsymbol{F}$ 的“方向余弦”。于是，单位矢量 $\boldsymbol{e}_F$就等于

$$\boldsymbol{e}_F = \cos\theta_x \boldsymbol{i} + \cos\theta_y \boldsymbol{j} + \cos\theta_z \boldsymbol{k} \tag{4.8}$$

也就是

$$\boldsymbol{e}_F = \left(\frac{F_x\boldsymbol{i} + F_y\boldsymbol{j} + F_z\boldsymbol{k}}{\sqrt{F_x^2 + F_y^2 + F_z^2}}\right) \tag{4.9}$$

方向角 θ_x、θ_y和 θ_z则分别为

$$\theta_x = \arccos\left(\frac{2.67}{10}\right) = 74.5°$$

$$\theta_y = \arccos\left(\frac{8.02}{10}\right) = 36.7°$$

$$\theta_z = \arccos\left(\frac{5.34}{10}\right) = 57.7°$$

将各个分量加起来就构成了矢量，即

$$\boldsymbol{A} = A_x\boldsymbol{i} + A_y\boldsymbol{j} + A_z\boldsymbol{k}$$

$$\boldsymbol{B} = B_x\boldsymbol{i} + B_y\boldsymbol{j} + B_z\boldsymbol{k}$$

$$\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B} = (A_x + B_x)\boldsymbol{i} + (A_y + B_y)\boldsymbol{j} + (A_z + B_z)\boldsymbol{k}$$

一般，几个力相加可以形成合力 $\boldsymbol{R}$，即

$$\boldsymbol{R} = \sum F_x\boldsymbol{i} + \sum F_y\boldsymbol{j} + \sum F_z\boldsymbol{k} \tag{4.10}$$

矢量的相减同样也等于各个分量的相减。

矢量的乘法有两种不同的运算，即点积和叉积。两个矢量 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的点积（即标量积）产生一个标量，即

$$\boldsymbol{A} \cdot \boldsymbol{B} = AB\cos\theta \tag{4.11}$$

其中 θ 是两个矢量之间的夹角。对于正交坐标系，各个坐标轴之间的夹角均为 90°，因此

$$\boldsymbol{i} \cdot \boldsymbol{i} = \boldsymbol{j} \cdot \boldsymbol{j} = \boldsymbol{k} \cdot \boldsymbol{k} = 1$$

$$\boldsymbol{i} \cdot \boldsymbol{j} = \boldsymbol{j} \cdot \boldsymbol{k} = \boldsymbol{k} \cdot \boldsymbol{i} = \cdots = 0 \tag{4.12}$$

例如

$$\boldsymbol{A} = 3\boldsymbol{i} + 2\boldsymbol{j} + \boldsymbol{k}\ \text{ft}$$

$$\boldsymbol{B} = -2\boldsymbol{i} + 3\boldsymbol{j} + 10\boldsymbol{k}\ \text{lb}$$

$$\boldsymbol{A} \cdot \boldsymbol{B} = 3(-2) + 2(3) + 1(10) = 10\text{ft lb}$$

注意，点积是可交换的，即 $\boldsymbol{A} \cdot \boldsymbol{B} \equiv \boldsymbol{B} \cdot \boldsymbol{A}$。

点积 $\boldsymbol{A} \cdot \boldsymbol{B}$ 的物理意义是 $\boldsymbol{A}$ 在 $\boldsymbol{B}$ 上的投影，也等于 $\boldsymbol{B}$ 在 $\boldsymbol{A}$ 上的投影。例如，“功”的定义就涉及作用力在物体运动方向上的投影。图 4.3 中左图所示就是力矢量 $\boldsymbol{F}$ 与运动矢量 $\boldsymbol{d}$ 的点积，力 $\boldsymbol{F}$ 所做的功等于 $\boldsymbol{F} \cdot \boldsymbol{d} \equiv Fd\cos\theta$，$\boldsymbol{F}$ 与 $\boldsymbol{d}$ 的点积等于 $\boldsymbol{d}$ 方

向上的 $\boldsymbol{F}$ 分量与 $\boldsymbol{d}$ 的乘积。

力矩表示力相对于某个点或者某个轴的转动趋势，两个矢量的叉积（即矢量积）产生一个方向与转轴一致的新矢量。例如，图 4.3 右图中的力 $\boldsymbol{F}$ 在 $x—y$ 平面上作用于某个物体，$\boldsymbol{F}$ 的作用点与物体坐标中心 O 相距一段距离，矢量 r 从 O 点指向 $\boldsymbol{F}$ 的作用线，叉积 $\boldsymbol{r}\times\boldsymbol{F}$ 的方向与 z 轴相同，就是物体转轴的方向。如果 $\boldsymbol{F}$ 和 $\boldsymbol{r}$ 是三维矢量，即存在 $\boldsymbol{k}$ 分量，那么，它们的叉积也会产生沿 x 轴和 y 轴的转动分量。$\boldsymbol{r}$ 与 $\boldsymbol{F}$ 的力矩 $\boldsymbol{M}$ 记为

$$\boldsymbol{M}=M_x\boldsymbol{i}+M_y\boldsymbol{j}+M_z\boldsymbol{k} \tag{4.13}$$

式中 M_x、M_y、M_z分别是物体绕 x、y、z 轴转动的分量。

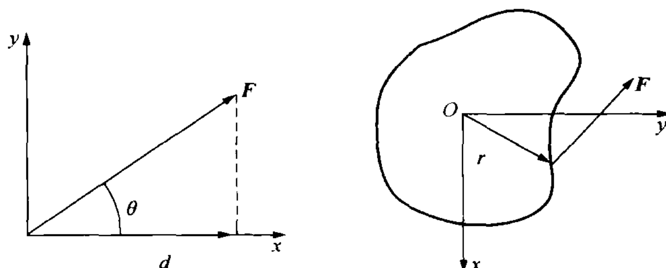


图 4.3 左：矢量 $\boldsymbol{F}$ 与矢量 $\boldsymbol{d}$ 的点积等于 $\boldsymbol{F}$ 在 $\boldsymbol{d}$ 上的投影；右：矢量 $\boldsymbol{r}$ 与矢量 $\boldsymbol{F}$ 的叉积仍然是一个矢量，其方向是转轴方向，也就是垂直于纸面向外的 z 轴方向

矢量的叉积可以通过各个分量的叉积之和来计算，例如

$$\begin{aligned}\boldsymbol{A}\times\boldsymbol{B}=&3(-2)\boldsymbol{i}\times\boldsymbol{i}+3(3)\boldsymbol{i}\times\boldsymbol{j}+3(10)\boldsymbol{i}\times\boldsymbol{k}+\\&2(-2)\boldsymbol{j}\times\boldsymbol{i}+2(3)\boldsymbol{j}\times\boldsymbol{j}+2(10)\boldsymbol{j}\times\boldsymbol{k}+\\&1(-2)\boldsymbol{k}\times\boldsymbol{i}+1(3)\boldsymbol{k}\times\boldsymbol{j}+1(10)\boldsymbol{k}\times\boldsymbol{k}\end{aligned}$$

其大小为$|\boldsymbol{A}\times\boldsymbol{B}|=AB\sin\theta$，$\theta$ 是 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 之间的夹角，因此，对于正交坐标系，所有相同坐标轴上分量的叉积都等于 0，并且 $\boldsymbol{i}\times\boldsymbol{j}=\boldsymbol{k}$，$\boldsymbol{j}\times\boldsymbol{k}=\boldsymbol{i}$，$\boldsymbol{k}\times\boldsymbol{i}=\boldsymbol{j}$，$\boldsymbol{i}\times\boldsymbol{k}=-\boldsymbol{j}$，于是上式就成为

$$\begin{aligned}\boldsymbol{A}\times\boldsymbol{B}&=9\boldsymbol{k}-30\boldsymbol{j}+4\boldsymbol{k}+20\boldsymbol{i}-2\boldsymbol{j}-3\boldsymbol{i}\\&=17\boldsymbol{i}-32\boldsymbol{j}+13\boldsymbol{k}\ \text{lb ft}\end{aligned}$$

注意，叉积是不可交换的，即 $\boldsymbol{A}\times\boldsymbol{B}\neq\boldsymbol{B}\times\boldsymbol{A}$。

矢量的叉积通常用行列式来计算，上述例子的 $\boldsymbol{A}\times\boldsymbol{B}$ 可以写成如下行列式形式

$$\begin{aligned}\boldsymbol{A}\times\boldsymbol{B}&=\begin{vmatrix}\boldsymbol{i}&\boldsymbol{j}&\boldsymbol{k}\\A_x&A_y&A_z\\B_x&B_y&B_z\end{vmatrix}\\&=\begin{vmatrix}\boldsymbol{i}&\boldsymbol{j}&\boldsymbol{k}\\3&2&1\\-2&3&10\end{vmatrix}\\&=\boldsymbol{i}[(2)(10)-(1)(3)]-\boldsymbol{j}[(3)(10)-(1)(-2)]+\boldsymbol{k}[(3)(3)-(2)(-2)]\\&=\boldsymbol{i}(20-3)-\boldsymbol{j}(30+2)+\boldsymbol{k}(9+4)\end{aligned}$$

$$=17\boldsymbol{i}-32\boldsymbol{j}+13\boldsymbol{k}\ \text{lb ft} \tag{4.14}$$

例 4.1　图 4.4 所示的矢量 $\boldsymbol{F}$ 的大小为 10kN[⊖]，其方向如虚线所示。(a) 请写出矢量 $\boldsymbol{F}$ 的表达式；(b) 矢量 $\boldsymbol{F}$ 在 $x-z$ 平面上的分量为多少？(c) 矢量 $\boldsymbol{F}$ 对于原点（0，0，0）所产生的力矩是多少？

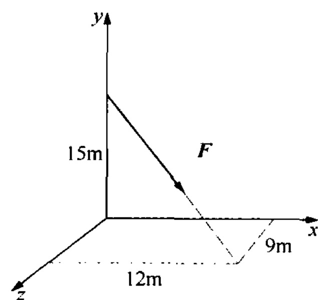


图 4.4　大小为 10kN 的力矢量 $\boldsymbol{F}$

解：下面用 MATLAB 指令求解本例题，其中，提示符"≫"之后是键入的内容，百分号"%"之后是注释，MATLAB 不执行注释中的内容。不带提示符"≫"的行是 MATLAB 的输出结果。为了节省空间，某些输出被省略。MATLAB 运行过程如下：

```
≫ %(a) First write the direction vector d that points along F
≫ % as a 1D array:
≫ d = [12 -15 9]

d = 12 -15 9
≫ % Now write the unit vector of F, giving its direction:
≫ unit_vector = d/norm (d)

unit_vector = 0.5657     -0.7071     0.4243

≫ % F consists of the magnitude 10 kN times this unit vector
≫ F = 10*unit_vector

F = 5.6569     -7.0711     4.2426

≫ % Or, more directly
≫ F = 10*(d/norm(d))

F = 5.6569     -7.0711     4.2426

≫ % (b) First write the vector r_xz that points in the xz plane:
≫ r_xz = [12 0 9]
r_xz = 12   0   9

≫ % The dot product is given by the sum of all the term by term
≫ % multiplications of elements of vectors F and r_xz
≫ F_dot_r_xz = sum(F.*r_xz)
≫ % or simply, dot(F,r_xz)

F_dot_r_xz = 106.0660
```

⊖ "kN" 是单位"千牛顿"。——译者注

```
≫ % (c) Cross F with a vector that points from the origin to F.
≫ % The cross product is given by the cross function
≫ r_xz_cross_F = cross(r_xz,F)

r_xz_cross_F = 63.6396     0     -84.8528

≫ % Note that the cross product is not commutative
≫ cross(F,r_xz)
ans = -63.6396     0     84.8528

≫ % Vectors are added and subtracted in MATLAB using the + and -
≫ % operations, respectively.
```
■

4.2.2 坐标变换

三维空间的方向余弦

研究人体运动时，经常需要把人体或者人体的某个肢体段从一个坐标系转换到另一个坐标系。例如，人体上粘贴的标记物所定义的坐标系是一个活动坐标系，它所定义的坐标点要变换到实验室空间的固定坐标系下。这种三维坐标变换可以通过方向余弦来计算，简介如下。

如图 4.5 所示，假设矢量 **A** 在大写字母表示的坐标系 XYZ 下用单位矢量 **I**、**J**、**K** 表示为

$$\boldsymbol{A}=A_x\boldsymbol{I}+A_y\boldsymbol{J}+A_z\boldsymbol{k} \tag{4.15}$$

而单位矢量 **I**、**J**、**K** 在小写字母表示的 xyz 坐标系中可以用单位矢量 **i**、**j**、**k** 表示为

$$\boldsymbol{I}=\cos\theta_{xX}\boldsymbol{i}+\cos\theta_{yX}\boldsymbol{j}+\cos\theta_{zX}\boldsymbol{k} \tag{4.16}$$

$$\boldsymbol{J}=\cos\theta_{xY}\boldsymbol{i}+\cos\theta_{yY}\boldsymbol{j}+\cos\theta_{zY}\boldsymbol{k} \tag{4.17}$$

$$\boldsymbol{K}=\cos\theta_{xZ}\boldsymbol{i}+\cos\theta_{yZ}\boldsymbol{j}+\cos\theta_{zZ}\boldsymbol{k} \tag{4.18}$$

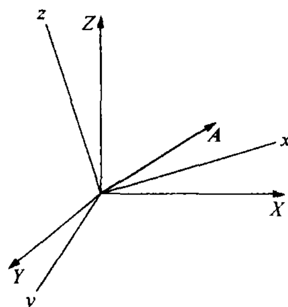


图 4.5 在 XYZ 坐标系中测得的矢量 **A** 可以通过式（4.20）的 9 个方向余弦变换到 xyz 坐标系中

式中，θ_{xX}是 **i** 与 **I** 之间的夹角，其他夹角类推。

将式（4.16）～式（4.18）代入式（4.15），可得

$$\begin{aligned}\boldsymbol{A}=&A_x\ [\cos\theta_{xX}\boldsymbol{i}+\cos\theta_{yX}\boldsymbol{j}+\cos\theta_{zX}\boldsymbol{k}]\ +\\&A_y\ [\cos\theta_{xY}\boldsymbol{i}+\cos\theta_{yY}\boldsymbol{j}+\cos\theta_{zY}\boldsymbol{k}]\ +\\&A_z\ [\cos\theta_{xZ}\boldsymbol{i}+\cos\theta_{yZ}\boldsymbol{j}+\cos\theta_{zZ}\boldsymbol{k}]\end{aligned} \tag{4.19}$$

也就是

$$\begin{aligned}\boldsymbol{A}=&(A_x\cos\theta_{xX}+A_y\cos\theta_{xY}+A_z\cos\theta_{xZ})\ \boldsymbol{i}+\\&(A_x\cos\theta_{yX}+A_y\cos\theta_{yY}+A_z\cos\theta_{yZ})\ \boldsymbol{j}+\\&(A_x\cos\theta_{zX}+A_y\cos\theta_{zY}+A_z\cos\theta_{zZ})\ \boldsymbol{k}\end{aligned} \tag{4.20}$$

于是，矢量 **A** 既可以用 **I**、**J**、**K** 表示，也可以用 **i**、**j**、**k** 表示。

Euler 角

人体在某个正交坐标系下的坐标可以通过欧拉角变换矩阵转换到另一个正交坐标系下。例如，某个坐标系由病人骨盆部位的标记物定义，而另一个坐标系则由病人大腿处的标记物定义，那么，经过围绕各个原始坐标轴的一系列旋转之后，就可以将一个坐标系下的点变换到另一个坐标系下。图 4.6 所示就是 xyz 坐标系按照 y 轴—x 轴—z 轴的顺序依次完成旋转的变换过程。首先，将 xyz 坐标系绕 y 轴旋转（图 4.6 的左图），将单位矢量 $\boldsymbol{i}$、$\boldsymbol{j}$、$\boldsymbol{k}$ 转变为单位矢量 $\boldsymbol{i}'$、$\boldsymbol{j}'$、$\boldsymbol{k}'$，变换方程是

$$\boldsymbol{i}'=\cos\theta_y\boldsymbol{i}-\sin\theta_y\boldsymbol{k} \tag{4.21}$$

$$\boldsymbol{j}'=\boldsymbol{j} \tag{4.22}$$

$$\boldsymbol{k}'=\sin\theta_y\boldsymbol{i}+\cos\theta_y\boldsymbol{k} \tag{4.23}$$

然后，再将这个新坐标系绕 x 轴旋转（图 4.6 的中图），变为

$$\boldsymbol{i}''=\boldsymbol{i}' \tag{4.24}$$

$$\boldsymbol{j}''=\cos\theta_x\boldsymbol{j}'+\sin\theta_x\boldsymbol{k}' \tag{4.25}$$

$$\boldsymbol{k}''=-\sin\theta_x\boldsymbol{j}'+\cos\theta_x\boldsymbol{k}' \tag{4.26}$$

最后，再绕 z 轴旋转（图 4.6 的右图），得到

$$\boldsymbol{i}'''=\cos\theta_z\boldsymbol{i}''+\sin\theta_z\boldsymbol{j}'' \tag{4.27}$$

$$\boldsymbol{j}'''=-\sin\theta_z\boldsymbol{i}''+\cos\theta_z\boldsymbol{j}'' \tag{4.28}$$

$$\boldsymbol{k}'''=\boldsymbol{k}'' \tag{4.29}$$

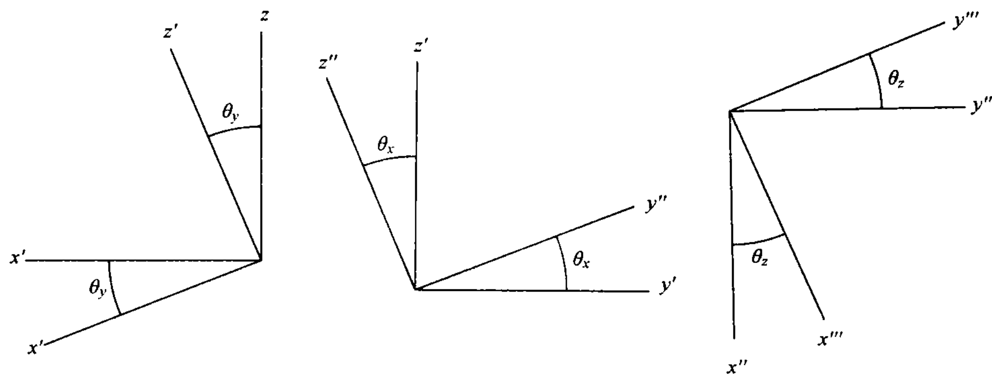


图 4.6　xyz 原始坐标系经过 3 次旋转之后转变成新坐标系 $x'''y'''z'''$。
左：绕 y 轴旋转；中：绕 x 轴旋转；右：绕 z 轴旋转

利用矩阵运算可以完成上述 3 次旋转，直接将单位矢量 $\boldsymbol{i}$、$\boldsymbol{j}$、$\boldsymbol{k}$ 变为单位矢量 $\boldsymbol{i}'''$、$\boldsymbol{j}'''$、$\boldsymbol{k}'''$，即

$$\begin{bmatrix} i''' \\ j''' \\ k''' \end{bmatrix}=\begin{bmatrix} \cos\theta_z & \sin\theta_z & 0 \\ -\sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & \sin\theta_x \\ 0 & -\sin\theta_x & \cos\theta_x \end{bmatrix}\begin{bmatrix} \cos\theta_y & 0 & -\sin\theta_y \\ 0 & 1 & 0 \\ \sin\theta_y & 0 & \cos\theta_y \end{bmatrix}\begin{bmatrix} i \\ j \\ k \end{bmatrix}$$

$$= \begin{bmatrix} \cos\theta_z & \sin\theta_z\cos\theta_x & \sin\theta_z\sin\theta_x \\ -\sin\theta_z & \cos\theta_z\cos\theta_x & \cos\theta_z\sin\theta_x \\ 0 & -\sin\theta_x & \cos\theta_x \end{bmatrix} \begin{bmatrix} \cos\theta_y & 0 & -\sin\theta_y \\ 0 & 1 & 0 \\ \sin\theta_y & 0 & \cos\theta_y \end{bmatrix} \begin{bmatrix} i \\ j \\ k \end{bmatrix} \tag{4.30}$$

$$\begin{bmatrix} i''' \\ j''' \\ k''' \end{bmatrix} = \begin{bmatrix} \cos\theta_z\cos\theta_y + \sin\theta_z\sin\theta_x\sin\theta_y & \sin\theta_z\cos\theta_x & -\cos\theta_z\sin\theta_y + \sin\theta_z\sin\theta_x\cos\theta_y \\ -\sin\theta_z\cos\theta_y + \cos\theta_z\sin\theta_x\sin\theta_y & \cos\theta_z\cos\theta_x & \sin\theta_z\sin\theta_y + \cos\theta_z\sin\theta_x\cos\theta_y \\ \cos\theta_x\sin\theta_y & -\sin\theta_x & \cos\theta_x\cos\theta_y \end{bmatrix} \begin{bmatrix} i \\ j \\ k \end{bmatrix} \tag{4.31}$$

如果坐标系的旋转角 θ_x、θ_y和θ_z已知，那么，xyz 坐标系下的坐标值就都可以转换到 $x'''y'''z'''$坐标系下。反之，如果这两个坐标系已知，那么，这些旋转角就都可以求得，即

$$k''' \cdot j = -\sin\theta_x \tag{4.32}$$

$$\theta_x = -\arcsin\ (k''' \cdot j)$$

$$k''' \cdot i = \cos\theta_x\sin\theta_y$$

$$\theta_y = \arcsin\left[\frac{k''' \cdot i}{\cos\theta_x}\right] \tag{4.33}$$

$$i''' \cdot j = \sin\theta_z\cos\theta_x$$

$$\theta_z = \arcsin\left[\frac{i''' \cdot j}{\cos\theta_x}\right] \tag{4.34}$$

例 4.2 请利用 MATLAB 的符号数学工具箱，编写 $y-x-z$ 旋转顺序的欧拉角变换矩阵。

解： MATLAB 程序如下：

```
% eulerangles.m
%
% Euler angles for y-x-z rotation sequence
% using MATLAB symbolic math toolbox
%
% x, y and z are thetax, thetay and thetaz, respectively
% First define them as symbolic variables
syms x y z

% Writing equations 4.21–23 as a matrix A

A = [ cos(y),   0,   -sin(y);
         0,       1,        0;
      sin(y),   0,    cos(y)]

% equations 4.24–26 as matrix B

B = [ 1,        0,          0;
      0,     cos(x),    sin(x);
      0,    -sin(x),    cos(x)]
```

```
% and equations 4.27–29 as matrix C
C = [ cos(z),   sin(z),   0;
     -sin(z),   cos(z),   0;
       0,         0,      1]

% The matrix equation 4.30 is created by multiplying matrices C, B
% and A

D = C*B*A
```

以上 eulerangles. m 文件运行之后得到的变换矩阵为

```
D =
[cos(z)*cos(y)+sin(z)*sin(x)*sin(y), sin(z)*cos(x), -cos(z)*sin(y)+sin(z)*sin(x)*cos(y)]
[-sin(z)*cos(y)+cos(z)*sin(x)*sin(y), cos(z)*cos(x), sin(z)*sin(y)+cos(z)*sin(x)*cos(y)]
[cos(x)*sin(y),                       -sin(x),                          cos(x)*cos(y)]
```

该输出结果与式（4. 31）相同。 ■

已知的数据情况不同，Euler 变换矩阵的应用方式也不同。例如，如果物体在两个坐标系中的坐标值都已知，那么，就可以计算 θ_x、θ_y和θ_z角，其中 $y-x-z$ 旋转顺序的计算公式就是式（4. 32）～式（4. 34）。反过来，如果物体的初始位置坐标和 θ_x、θ_y、θ_z角已知，也可以计算物体的最终位置坐标，下面举例说明。

例 4. 3 假设一架飞机先倾斜了 30°（即 θ_x），然后又旋转了 20°（即 θ_y），最后偏斜了 10°（即 θ_z），请编写一个 MATLAB 函数，计算这一系列旋转的欧拉角变换矩阵。

解： 由于计算机中三角函数的计算单位是弧度，下面先编写两个简单的 MATLAB 函数 cosd 和 sind，用于计算以 “°” 为单位的余弦和正弦三角函数值，即

```
function y = cosd(x)
%COSD(X) cosines of the elements of X measured in degrees.
y = cos(pi*x/180);

function y = sind(x)
%SIND(X) sines of the elements of X measured in degrees.
y = sin(pi*x/180);
```

然后编写 $x-y-z$ 旋转顺序的变换矩阵，即

```
function D = eulangle (thetax, thetay, thetaz)
%EULANGLE matrix of rotations by Euler's angles.
%   EULANGLE(thetax, thetay, thetaz) yields the matrix of
%   rotation of a system of coordinates by Euler's
%   angles thetax, thetay and thetaz, measured in degrees.

% Now the first rotation is about the x axis, so we use eqs. 4.24–26

A = [ 1        0                  0
      0        cosd(thetax)       sind(thetax)
      0        -sind(thetax)      cosd(thetax) ];
```

```
% Next is the y axis rotation (Eqs. 4.21-23)

B = [ cosd(thetay)       0        -sind(thetay)
      0                  1        0
      sind(thetay)       0        cosd(thetay) ];

% Finally, the z axis rotation (Eqs. 4.27-29)

C = [  cosd(thetaz)      sind(thetaz)     0
       -sind(thetaz)     cosd(thetaz)     0
       0                 0                1   ];

% Multiplying rotation matrices C, B and A as in Eq. 4.30 gives the solution:

D=C*B*A;
```

运行该 MATLAB 函数，得到如下变换矩阵的输出结果：

```
>> eulangle(30,20,10)
ans =
    0.9254    0.3188   -0.2049
   -0.1632    0.8232    0.5438
    0.3420   -0.4698    0.8138
```

该矩阵可以将初始坐标系下任意的位点转换到飞机动作之后所达到的位置。 ■

4.2.3 静力学平衡

描述静力学平衡结构的牛顿运动方程可以用如下矢量方程表示

$$\sum F=0 \tag{4.35}$$

$$\sum M=0 \tag{4.36}$$

这两个方程在生物系统中的应用方法与标准力学结构中相同。首先需要画出研究对象分离体的受力图，即在各个受力点上标出外加的负荷和作用力。如第 3 章的图 3.33 所示，人工关节可以用铰链关节和球窝关节等理想化的模型来模拟。

例 4.4 图 4.7 中的左图是 Russell 牵引法的示意图，该装置用于给骨折的股骨施加一个轴向拉力。(a) 在牵拉绳索的悬空端需要施加多少重量 w 才能使腿部保持静态平衡？(b) 请计算图示条件下施加在小腿上的平均拉力。

解： 该系统的分离体受力图如图 4.7 右图所示。假设滑轮无摩擦力，且其半径很小，绳索的拉力 T 可以看做始终是常数，那么，利用式 (4.35) 可得

$$\boldsymbol{F}_1+\boldsymbol{F}_2+\boldsymbol{F}_3+\boldsymbol{F}_{\text{femur}}-mg\boldsymbol{j}=0$$

将每个作用力写成矢量形式，即

$$\boldsymbol{F}_1=-F_1\boldsymbol{i}=-T\boldsymbol{i}$$

$$\begin{aligned}\boldsymbol{F}_2&=(-F_2\cos 30°)\boldsymbol{i}+(F_2\sin 30°)\boldsymbol{j}\\&=(-T\cos 30°)\boldsymbol{i}+(T\sin 30°)\boldsymbol{j}\end{aligned}$$

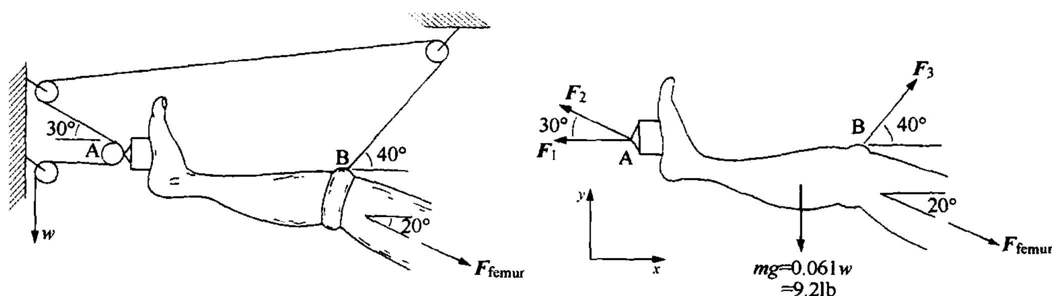


图 4.7　左：临床上用于给腿部加负荷的 Russell 牵引装置；右：牵引时腿部的受力图（此图根据 Davis（1986）著作第 206～207 页的图 6.14 和图 6.15 修改）

$$\boldsymbol{F}_3=(F_3\cos 40^\circ)\boldsymbol{i}+(F_3\sin 40^\circ)\boldsymbol{j}$$
$$=(T\cos 40^\circ)\boldsymbol{i}+(T\sin 40^\circ)\boldsymbol{j}$$
$$\boldsymbol{F}_{femur}=(F_{femur}\cos 20^\circ)\boldsymbol{i}-(F_{femur}\sin 20^\circ)\boldsymbol{j}$$

不计大腿的重量。查表 4.1 可知，脚和小腿的重量等于总体重（150 lb）乘以 0.061，即

$$mg\boldsymbol{j}=(0.061)(150\boldsymbol{j}\text{lb})=9.2\boldsymbol{j}\text{lb}$$

x 分量之和为

$$-T-T\cos30^\circ+T\cos40^\circ+F_{femur}\cos20^\circ=0$$

y 分量之和为

$$T\sin30^\circ+T\sin40^\circ-F_{femur}\sin20^\circ-mg=0$$

将以上两个方程联列求解，可得 T 和轴向拉力 F_{femur} 为：

$$T=12.4\text{ lb}$$
$$F_{femur}=14.5\text{ lb}$$

T 就等于需要外加的重量 w。■

例 4.5　某人体重为 160 lb，如图 4.8 的上图所示，其手掌上托着 10 lb 的重球，肘部呈 90°角。(a) 此时肱二头肌需要产生多大的力才能使前臂保持平衡？(b) 前臂施加在肱骨上的力为多大？

解：图 4.8 的下图是该系统的分离体受力图，由于未知数比例题 4.4 多，因此，式（4.35）和式（4.36）都要用上。关于肘部 O 点的力矩之和的平衡方程 $\boldsymbol{\Sigma M}=0$ 可以写为

$$-\boldsymbol{r}_{OE}\times(-\boldsymbol{F}_A)+\boldsymbol{r}_{OB}\times(-10\text{ lb})\boldsymbol{j}+\boldsymbol{r}_{OP}\times(-3.5\text{ lb})\boldsymbol{j}=0$$
$$(-2\text{in})\boldsymbol{i}\times(-F_A)\boldsymbol{j}+(12\text{in})\boldsymbol{i}\times(-10\text{ lb})\boldsymbol{j}+(9.25\text{in})\boldsymbol{i}\times(-3.5\text{ lb})\boldsymbol{j}=0$$
$$(2\text{in})F_A\boldsymbol{k}-(120\text{ lb in})\boldsymbol{k}-(32.4\text{ lb in})\boldsymbol{k}=0$$

求解以上最后一个方程可得肘部受到的垂直力 F_A 为

$$F_A=76.2\text{ lb}$$

为了求得肘部受到的水平力 F_C 和肱二头肌所产生的力 F_B，要再加上平衡方程 $\boldsymbol{\Sigma F}=0$，即

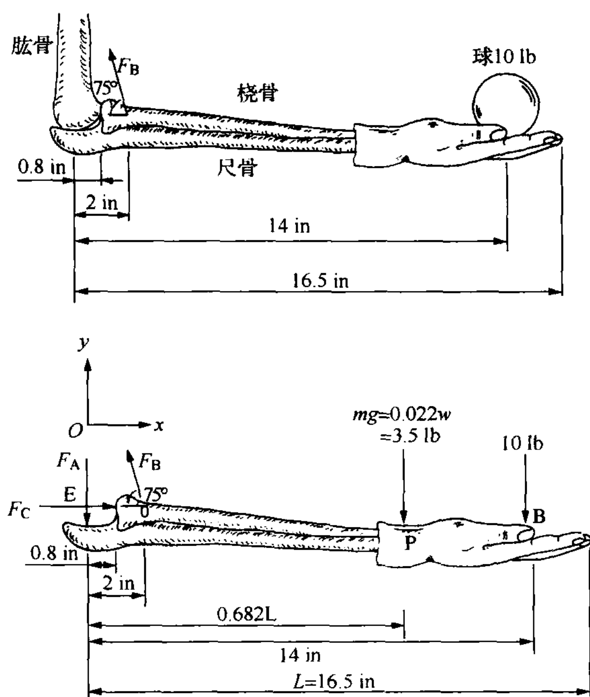


图 4.8 上图：肘部静止并呈 90°角，手掌上托着 10lb 的重球。下图：前臂的受力图（此图根据 Davis（1986）著作第 208～209 页的图 6.16 和图 6.17 修改）

$$F_C\boldsymbol{i}-F_A\boldsymbol{j}+(-F_B\cos75°\boldsymbol{i}+F_B\sin75°\boldsymbol{j})-10\ \text{lb}\ \boldsymbol{j}-3.5\ \text{lb}\ \boldsymbol{j}=0$$

分别求 x 和 y 分量之和，可得

$$F_C-F_B\cos(75°)=0$$

$$-F_A+F_B\sin(75°)-10\ \text{lb}-3.5\ \text{lb}=0$$

将 $F_A=76.2$ lb 代入，联列求解这两个方程，可得

$$F_B=92.9\ \text{lb}$$

$$F_C=24.1\ \text{lb}$$

■

例 4.6 图 4.9 所示测力平台的 4 个角上分别安装了 1 个传感器。假设这些传感器测得的垂直力分别为 F_1、F_2、F_3 和 F_4，如果平台是边长为 l 的正方形，且 F_1～F_4 已知，请分别写出合力 R 的 x 坐标和 y 坐标表达式。

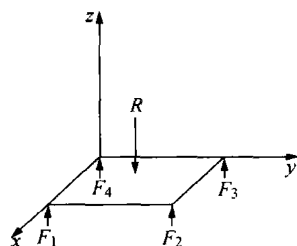


图 4.9 边长为 l 的正方形测力平台受到合力 R 的作用，其 4 个角上各安装了 1 个传感器，分别测量垂直作用力 F_1～F_4

解： 合力 R 的大小可以由 z 方向作用力之和求得，即

$$\sum F_z=0$$

$$F_1+F_2+F_3+F_4-R=0$$

$$R=F_1+F_2+F_3+F_4$$

测力平台保持水平，因此，绕 x 轴和 y 轴的力矩之和必须都为 0。绕 x 轴的力矩之和为

$$\sum M_x=0$$

$$F_2l+F_3l-Ry=0$$

$$y=\frac{(F_2+F_3)l}{R}$$

同理，绕 y 轴的力矩之和为

$$\sum M_y=0$$

$$F_1l+F_2l-Rx=0$$

$$x=\frac{(F_1+F_2)l}{R}$$

这就是合力 R 的 x 坐标和 y 坐标计算公式。 ■

4.2.4 转动惯量

物体的质量会阻碍其直线运动，而其转动惯量则会阻碍其转动。物体转动时所受到的阻力用如下转动惯量来描述：

$$I=\int_m r^2\mathrm{d}m \tag{4.37}$$

式中　m——物体的质量；

r——质元到转轴的距离；

$\mathrm{d}m$——质元，也可以写成 $\rho\mathrm{d}V$。

如果物体密度 ρ 为常数，则其转动惯量可以用物体体积 V 的积分来计算，即

$$I=\rho\int_V r^2\mathrm{d}V \tag{4.38}$$

根据这个通式，可以分别写出如下分别关于 x 轴、y 轴和 z 轴旋转的转动惯量：

$$\begin{aligned} I_{xx}&=\int_V(y^2+z^2)\rho\mathrm{d}V \\ I_{yy}&=\int_V(x^2+z^2)\rho\mathrm{d}V \\ I_{zz}&=\int_V(x^2+y^2)\rho\mathrm{d}V \end{aligned} \tag{4.39}$$

假设旋转半径 k 是转轴与物体质心之间的距离，那么，如果用质心来表示物体，其转动惯量变为

$$I=mk^2 \tag{4.40}$$

式中　m——物体的质量。

平行轴定理指出，物体对任一轴的转动惯量 I 等于物体对通过质心并与该轴平

行的轴的转动惯量 I_{cm} 加上物体质量与两轴之间距离 d 二次方的乘积，即

$$I=I_{cm}+md^2 \tag{4.41}$$

表 4.1 列出了人体各个肢体段的数据。

表 4.1　人体结构的数据

肢体段	定义	肢体段重量与体重之比	质心与肢体段长度之比 近端	远端	旋转半径与肢体段长度之比 近端	远端
手	腕关节至中指尖	0.006	0.506	0.494	0.587	0.577
前臂	肘关节至尺骨茎突	0.016	0.430	0.570	0.526	0.647
上臂	盂肱关节至肘关节	0.028	0.436	0.564	0.542	0.645
前臂与手	肘关节至中指尖	0.022	0.682	0.318	0.827	0.565
整条手臂	盂肱关节至尺骨茎突	0.050	0.530	0.470	0.645	0.596
脚	外踝至第二跖骨头	0.0145	0.50	0.50	0.690	0.690
小腿	股骨髁至内踝	0.0465	0.433	0.567	0.528	0.643
大腿	股骨大转子至股骨髁	0.100	0.433	0.567	0.540	0.653
脚与小腿	股骨髁至第二跖骨头	0.061	0.606	0.394	0.735	0.572
整条腿	股骨大转子至内踝	0.161	0.447	0.553	0.560	0.650
头与颈	C7－T1 与第一肋骨至耳道	0.081	1.000		1.116	
肩	胸骨锁骨关节至盂肱关节		0.712	0.288		
胸	C7－T1 至 T12－L1 和横隔膜	0.216	0.82	0.18		
腹	T12－L1 至 L4－L5	0.139	0.44	0.56		
臀	L4－L5 至股骨大转子	0.142	0.105	0.895		
胸与腹	C7－T1 至 L4－L5	0.355	0.63	0.37		
腹与臀	T12－L1 至股骨大转子	0.281	0.27	0.73		
躯干	股骨大转子至盂肱关节	0.497	0.50	0.50		
躯干、头、颈		0.578	0.66	0.34	0.830	0.607
头、手臂、躯干		0.678	0.626	0.374	0.798	0.621

此表根据 Winter（1990）著作第 56～57 页的表 3.1 修改。

例 4.7　某人体重 150lb，大腿长为 17in。请计算其大腿相对于质心的转动惯量（用 SI 国际单位制）。

解： 将大腿长度换算成 SI 单位制，即

$$l_{thigh}=17\text{in}=0.432\text{m}$$

表 4.1 列出了人体各肢体段重量与体重之比。将该人体体重换算成 SI 单位制为

$$m_{body}=(150lb)(0.454kg/lb)=68.1kg$$

于是，大腿重量为

$$m_{thigh}=(0.100)(68.1kg)=6.81kg$$

表 4.1 还列出了人体各肢体段相对于其长度的质心比值和旋转半径比值，并分别给出了近端和远端的长度比值。注意，大腿的“近端”是指靠近髋关节的一端，而“远端”则是指靠近膝关节的一端。因此，大腿的“近端”长度是指大腿质心与髋关节之间的距离，而“远端”长度则是指大腿质心与膝关节之间的距离。于是，大腿相对于髋关节的转动惯量为

$$I_{thigh/hip}=mk^2=(6.81kg)[(0.540)(0.432m)]^2=0.371kg\cdot m^2$$

由式（4.41）表示的平行轴定理可知，大腿相对于髋关节的转动惯量与大腿相对于其质心的转动惯量之间具有如下关系：

$$I_{thigh/hip}=I_{thigh/cm}+md^2$$

因此

$$I_{thigh/cm}=I_{thigh/hip}-md^2$$

距离 d 可以从大腿质心的近端长度数据求得，即

$$d=(0.432m)(0.433)=0.187m$$

于是，大腿相对于其质心的转动惯量为

$$I_{thigh/cm}=0.371kg\cdot m^2-(6.81kg)(0.187m)^2=0.133kg\cdot m^2$$ ■

4.2.5 运动方程

物体直线运动和转动的动力学可以用运动矢量方程来描述。

牛顿运动方程

牛顿第二定律表明了净作用力 $\boldsymbol{F}$ 及其引起的物体直线运动之间的关系，即

$$\boldsymbol{F}=m\boldsymbol{a} \tag{4.42}$$

式中 $\boldsymbol{a}$——物体质心直线运动的加速度。

对于转动有

$$\boldsymbol{M}=\boldsymbol{I\alpha} \tag{4.43}$$

式中 $\boldsymbol{I\alpha}$——物体的角动量。该式表明物体的角动量等于外界作用于物体的净力矩。

以上两个矢量运动方程一般都可以写成 x、y 和 z 分量组成的 6 个联立方程。

Euler 运动方程

牛顿运动方程描述了物体质心的运动，更一般化，欧拉运动方程则描述了刚体相对于其质心的运动。在 xyz 坐标与物体的主轴一致的特殊情况下，有

$$\sum M_x = I_{xx}\alpha_x + (I_{zz} - I_{yy})\omega_y\omega_z \tag{4.44}$$

$$\sum M_y = I_{yy}\alpha_y + (I_{xx} - I_{zz})\omega_z\omega_x \tag{4.45}$$

$$\sum M_z = I_{zz}\alpha_z + (I_{yy} - I_{xx})\omega_x\omega_y \tag{4.46}$$

式中 M_i——净力矩；

I_{ii}——物体相对于主轴的转动惯量；

α_i——物体的角加速度；

ω_i——物体的角速度。

欧拉方程中角度的单位是弧度。本书不详细推导该方程，有关内容在动力学的书籍中都可以找到。本章4.6节将利用式（4.44）~式（4.46）计算人体关节所受到的力矩。

4.3 材料力学

运动学和动力学用于描述生物体的运动及其相关的作用力，而材料力学则用于定量分析组织的形变，研究整形外科手术中作用力的分布，并预测植入的整形假体的性能以及手术矫正可能达到的效果。材料力学的研究领域非常广泛，下面将举例说明几个具有代表性的应用。

整形外科使用的接骨板是一块不锈钢板，它可以将断裂的骨头连接起来。如图4.10所示，接骨板用316L型不锈钢制成，其横截面A为4.17mm×12mm的长方形，如果轴向施加500N的负荷力F，则产生的轴向应力σ（即力/面积）为

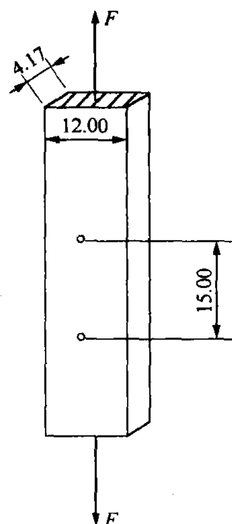


图4.10 用于固定骨折的接骨板，加了轴向负荷（图中尺寸的单位为mm）（摘自Burstein和Wright（1994）著作第104~108页，并有所改动）

$$\sigma=\frac{F}{A}=\frac{500\text{N}}{(4.17\times10^{-3}\text{m})(12\times10^{-3}\text{m})}=10\text{MPa} \tag{4.47}$$

与负荷力方向成45°角的平面上会产生最大剪切应力，其值τ_{max}为

$$\tau_{max}=\frac{F_{45^\circ}}{A_{45^\circ}}=\frac{(500\text{N})\cos45^\circ}{\left[\frac{(0.00417\text{m})(0.012\text{m})}{\cos45^\circ}\right]}=5\text{MPa} \tag{4.48}$$

该值等于0.5σ，符合材料力学的基本原理。如图4.10所示，在加负荷之前，钢板的长轴方向相距15.00mm处打了两个孔。加上500N负荷之后，两孔之间的距离增加了0.00075mm，则钢板的长度变化量Δl与原长度l之比就是应变ε

$$\varepsilon = \frac{\Delta l}{l} = \frac{0.00075\text{mm}}{15\text{mm}} = 50 \times 10^{-6} \tag{4.49}$$

常称为 50μ。此处μ表示 10^{-6}。

弹性模量 E 表示材料在受到外界拉力或者压力作用时的抗变形能力。线性弹性材料的 E 为常数，这种材料常称为胡克型材料（Hookean），其 σ 随 ε 变化的曲线是一条斜率恒定的直线，即

$$E = \frac{\sigma}{\varepsilon} \tag{4.50}$$

对于上述接骨板，有

$$E = \frac{10 \times 10^{6}\text{Pa}}{50 \times 10^{-6}} = 200\text{GPa}$$

金属和塑料等材料都只是在有限的负荷范围内才表现为线性，生物材料则具有更复杂的弹性性质。图 4.11 显示了在拉力作用下测得的骨组织的纵向和横向应力一应变曲线。先来看纵向应力一应变曲线，曲线上 0～7000μ 为弹性区，此时骨组织表现为纯粹的弹性体，$E \approx 12$GPa；当拉应力接近 90MPa 时，应力一应变曲线变成非线性，进入骨组织变形的塑性区；最后，当拉应力达到 120MPa 时，骨组织样本被破坏。这 3 个阶段的特性分别可以用材料的弹性模量（elastic modulus）、屈服应力（yield stress）和极限应力（ultimate stress）来表示，表 4.2 列出了常用整形材料（包括天然材料和人造材料）的这 3 个参数。

表 4.2 某些常用整形材料的屈服应力 σ_{yield}、极限应力 $\sigma_{ultimate}$ 和弹性模量 E

材 料	σ_{yield}/MPa	$\sigma_{ultimate}$/MPa	E/GPa
不锈钢	700	850	180
钴合金	490	700	200
钛合金	1100	1250	110
骨	85	120	18
PMMA		35	5
UHMWPE	14	27	1
膝盖韧带		58	

注：数据取自 Burstein 和 Wright（1994）著作第 122 页的表 4.2。⊖

图 4.11 还表明骨的纵向和横向弹性性质是不同的，骨是各向异性的。与纵向相比，骨在横向上要脆弱得多，硬度也要小得多，这从两个方向上的屈服应力、极限应力以及应力一应变曲线的斜率上的差别都可以看出。

图 4.12 表明在负荷的加载和卸载过程中骨所表现的弹性性质也不同，显示出弹性滞后现象。从热力学观点看，就是负荷加载过程中传入骨骼的能量与卸载过程中

⊖ PMMA 的化学名称是聚甲基丙烯酸甲酯，UHMWPE 是超高分子量聚乙烯。——译者注

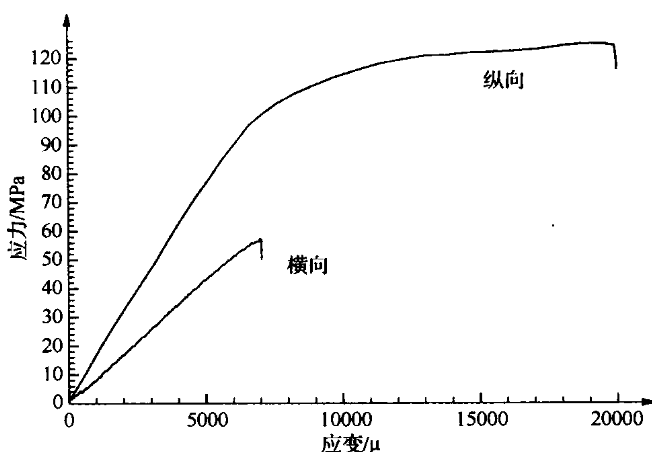


图 4.11 拉应力作用下骨的纵向和横向应力—应变曲线
（摘自 Burstein 和 Wright（1994）著作第 116 页的图 4.12，有所改动）

释放出来的能量不相等。并且，从曲线 A 到 B 再到 C，随着最大负荷的增加，这种能量差变得越来越大。“丢失”的能量损耗在骨骼内部摩擦引起的发热上，也损耗在大负荷引起的材料损伤上。

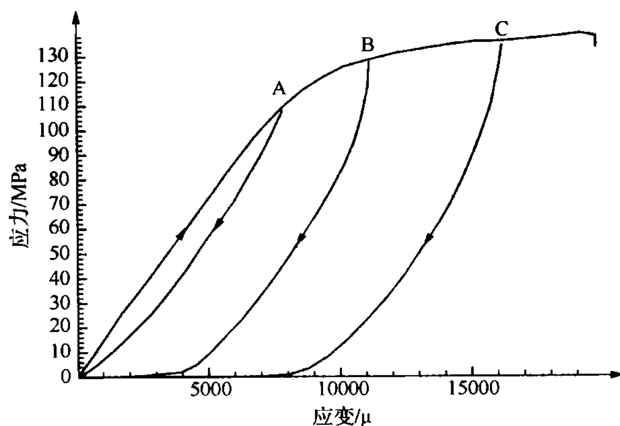


图 4.12 反复加载和卸载过程中的应力—应变曲线，即骨的弹性滞后现象
（摘自 Burstein 和 Wright（1994）著作第 119 页的图 4.15）

骨的各向异性很明显，受压时其极限应力为 200MPa，受拉时极限应力为 140MPa，而弯曲时的剪切极限应力只有 75MPa。剪切作用可以用剪切模量 G（又称刚性模量）来描述，它等于剪应力与剪应变之比。泊松比（Poisson's ratio）等于横向应变与纵向应变之比，用 ν 表示，即

$$\nu=\frac{\varepsilon_{横向}}{\varepsilon_{纵向}} \tag{4.51}$$

一般，$\nu\approx0.3$，表示纵向形变是横向形变的3倍。对于线性弹性材料，弹性模量E、剪切模量G和泊松比ν之间存在如下关系：

$$G=\frac{E}{2(1+\nu)} \tag{4.52}$$

生物材料的受损很难预料，其原因之一就是生理负荷的复杂性。例如，骨的抗拉能力比抗压能力要弱得多，滑雪时小腿胫骨在靴子顶端处容易骨折就是这个道理。由于脚被固定，滑雪者的前向冲量在靴子的顶端产生一个力矩，从而在小腿胫骨上形成三点弯曲。在这种弯曲状态下，胫骨前部受压，而后部受拉，很容易造成骨折。此时，小腿的三头肌（即腓肠肌与比目鱼肌）会在胫骨后侧产生高强度压应力，以减轻胫骨的受拉程度。

下面举例说明如何应用静力学和材料力学的方法解决生物力学问题。

例4.8　图4.13左图所示是用于固定股骨粗隆骨折的骨钉一骨板结构的植入体。假设人静止站立时臀部施加在植入体上的外力为400N。植入体的骨钉是长方形不锈钢板，截面尺寸为10mm×5mm，用几颗螺钉固定在股骨上，一部分沿x轴插入转子的头部。请问该植入体所承受的作用力、力矩、应力和应变各为多少？

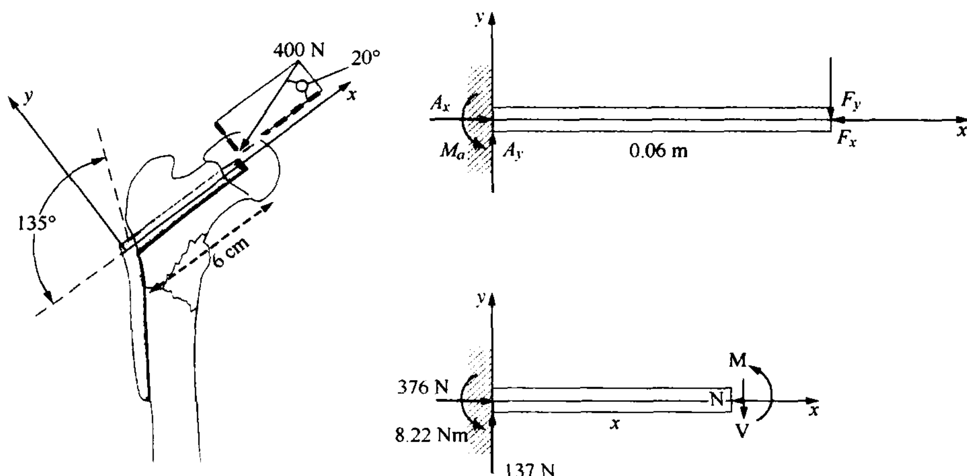


图4.13　左：用于股骨粗隆骨折修复的骨钉（取自Burstein和Wright（1994）著作第141页的图5.5）；右：上图为骨钉的受力图，下图为距左端支撑点x处的截面上的受力图

解：这是一个静力学问题，首先作受力图，画出各个受力点上受到的所有作用力及其力矩，以及反作用力及其力矩。由于骨折部位的不稳定性，骨钉一骨板植入体可能要承受整个400N负荷，因此，可以将骨钉模拟成一根长0.06m的悬梁，其所受负荷如图4.13右上所示。400N负荷可以分解为轴向F_x和横向F_y两个分量，即

$$F_x=400\text{ N}\quad\cos20^\circ=376\text{N}$$

$$F_y=400\text{ N}\quad\sin20^\circ=137\text{N}$$

轴向负荷分量产生压应力σ_x，根据式（4.47），有

$$\sigma_x = \frac{F_x}{A}$$

$$= \frac{376\ \text{N}}{(0.005\text{m})(0.01\text{m})} = 7.52\text{MPa}$$

该数值大约只有不锈钢屈服应力的1%（见表4.2）。轴向负荷引起的最大剪切应力为

$$\tau_{\max} = \frac{\sigma_x}{2} = 3.76\text{MPa}$$

其方向与长轴呈45°。根据不锈钢的弹性模量可以求得轴向应变，由于

$$E = \frac{\sigma}{\varepsilon} = \frac{F/A}{\Delta l/l}$$

所以

$$\varepsilon = \frac{F}{EA}$$

$$= \frac{376\text{N}}{180 \times 10^9 \text{Pa}(0.005\text{m})(0.01\text{m})} = 41.8 \times 10^{-6}$$

由此可以计算轴向变形为

$$\Delta l_{\text{axial}} = \varepsilon l = 2.51 \times 10^{-6}\,\text{m}$$

该变形很小，可以忽略不计。

横向负荷分量会引起悬梁弯曲，有关材料力学的书籍（如Roark1989年的著作）都有描述悬梁弯曲的方程式。假设图4.14左边所示的悬梁左端固定，右端受到一个向下的负荷，那么悬梁就会产生弯曲。此时梁的上面部分被拉伸，而下面部分则被压缩，中间既不受拉也不受压的转折位置称为中性轴（neutral axis）。对于厚度为h的对称长方形悬梁，中性轴位于中间，即图中c等于$h/2$。悬梁的抗弯曲特性可以用截面惯量（area moment of inertia）I来表示。如果长方形截面宽为b，厚为h，如图中右边所示，则$I = \frac{1}{12}bh^3$。

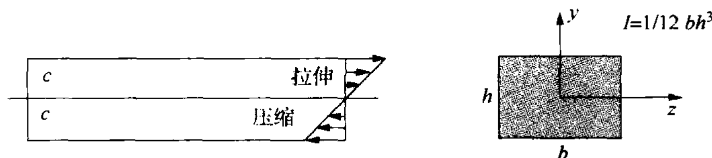


图4.14 左图：左端固定右端受到向下作用力的悬梁会产生弯曲，梁的上部分受到拉伸，而下部分则受到压缩，上下两部分的中间转折位置是中性轴；右图：宽为b厚为h的长方形截面悬梁的抗弯曲性可以用截面惯量I表示，此处$I = \frac{1}{12}bh^3$

悬梁末端的偏移量δy为

$$\delta y = \frac{Fx^2}{6EI}(3L - x) \tag{4.53}$$

式中 x——悬梁的轴向距离；

L——悬梁总长；

I——悬梁的截面惯量。

例如：

$$I=\frac{1}{12}(10\times10^{-3}\text{m})(5\times10^{-3}\text{m})^3=10.42\times10^{-9}\text{m}^4$$

悬梁最大的偏移发生在 $x=L$ 处，即

$$\begin{aligned}\delta y_{\max}&=\frac{FL^3}{3EI}\\&=\frac{137\text{N}(0.06\text{m})^3}{3(180\times10^9\text{N/m}^2)(10.42\times10^{-9}\text{m}^4)}\\&=5.26\times10^{-4}\text{m}=0.526\text{mm}\end{aligned} \tag{4.54}$$

该数值也可以忽略不计。

为了计算最大剪应力和弯应力，首先要求得最大剪切力 F 和弯矩 M。下面先列出悬梁整个受力图的静力学平衡方程，即

$$\sum F_x：A_x-376\text{N}=0$$

$$\sum F_y：A_y-137\text{N}=0$$

$$\sum M_A：M_a-137\text{N}(0.06\text{m})=0$$

求解该方程组，可得 $A_x=376\text{N}$，$A_y=137\text{N}$，$M_a=8.22\text{Nm}$。如果在左端固定点 A 右边的任意位置 x 处将悬梁截断，则悬梁左边部分的受力图如图 4.13 右下方所示。再对其应用静力学平衡方程则有

$$\sum F_x：376\text{N}-N=0$$

$$N(x)=376\text{N}$$

$$\sum F_y：137\text{N}-\text{V}=0$$

$$\text{V}(x)=137\text{N}$$

$$\sum M_A：8.22\text{Nm}-(137\text{N})(x\text{m})+M=0$$

$$M(x)=(137\text{Nm})x-8.22\text{Nm}$$

以下 MATLAB 程序可以显示这些方程的曲线图。如果轴向作用力和剪切力已知，那么弯矩曲线如图 4.15 所示。

```
% Use MATLAB to plot axial force, shear force, and bending moment diagrams
% for Example Problem 4.8

X = [0:0.01:0.06];
N = x.*0 + 376;
V = x.*0 + 137;
M = 137.*x - 8.22;

figure
subplot (3,1,1), plot(x,N,x,N, 'x')
```

```
xlabel ('x [m]')
ylabel('N [N]')
title ('Axial Force N')

subplot(3,1,2), plot(x,V,x,V,'x')
xlabel('x [m]')
ylabel('V [N]')
title('Shear Force V')
subplot(3,1,3), plot(x,M,x,M,'x')
xlabel('x [m]')
ylabel('M [N-m]')
title ('Bending Moment M')
```

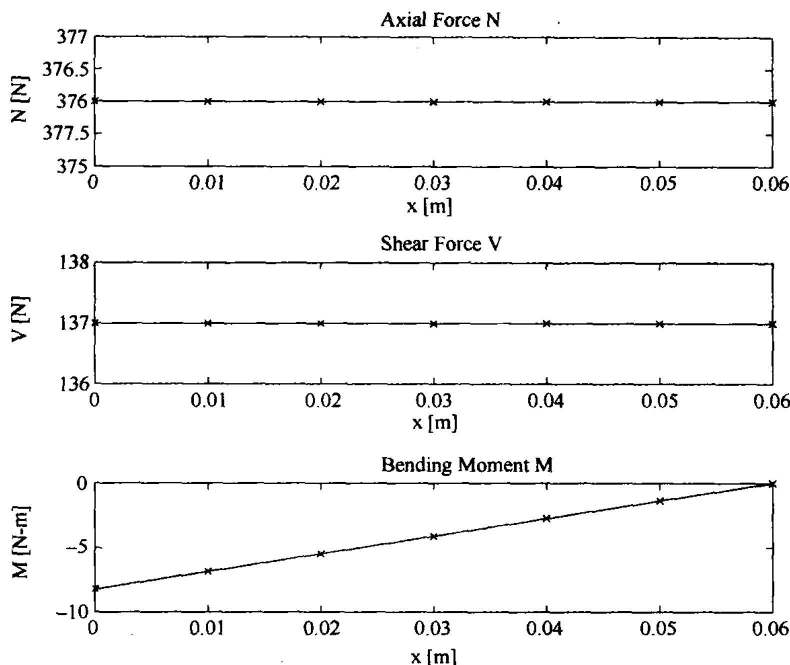


图 4.15 图 4.13 所示骨钉的轴向力 N、剪切力 V 和弯矩 M 随距离 x 变化的函数

最大弯应力计算公式为

$$\sigma_{b\max}=\frac{M_{\max}c}{I} \tag{4.55}$$

式中 c——最大应力处（即悬梁边缘）与中性轴的距离，此处即为 $h/2$。

于是，最大弯应力和最大剪应力分别为：

$$\sigma_{b\max}=\frac{-8.22\text{Nm}[0.5(5\times10^{-3}\text{m})]}{10.42\times10^{-9}\text{m}^4}=-197\text{MPa}$$

$$\tau_{b\max}=\frac{V_{\max}h^2}{8I} \tag{4.56}$$

$$=\frac{137\text{N}(5\times10^{-3}\text{m})^2}{8(10.42\times10^{-9}\text{m}^4)}=4.11\text{MPa}$$

这些应力都比不锈钢的屈服应力 $\sigma_{\text{yield}}=700\text{MPa}$ 小得多。■

4.4 粘弹性

胡克的固体弹性理论只有在很窄的负荷范围内才能正确描述材料的性质。例如，理想弹簧受到的作用力与弹簧长度的伸长量之间成正比关系，比例系数为常数 k。但是，对于实际的弹簧，在负荷很小或者很大的情况下，这种关系都不成立。还有，如果这种理想弹簧与一个运动物体相连，那么，由于忽略了阻尼，系统会产生纯粹的谐振，永远振荡下去，而实际上这种情况是不可能发生的。实际系统总是存在粘性，即阻尼性质，因此，系统的能量会通过弹簧的发热以及运动物体与空气之间的摩擦而逐渐损耗。

同理，生物材料都具有粘弹性。描述物体粘弹性的力学模型有很多，它们都有简单的本构方程。图 4.16 就显示了 3 种模型。其中，Maxwell 模型由弹簧和阻尼器串联构成；Voight 模型由弹簧和阻尼器并联构成；而 Kelvin 模型则由弹簧与阻尼器串联之后再与弹簧并联构成。每个模型都包含了阻尼器，阻尼器产生的作用力与伸缩量的导数成正比。这些模型所描述的应力和应变特性都会随时间变化。

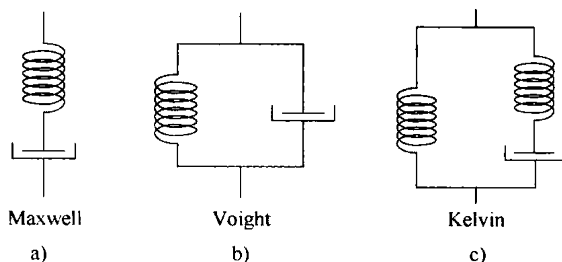


图 4.16　3 种简单的粘弹性模型
a) Maxwell 模型　b) Voight 模型
c) Kelvin 模型（即固体的标准线性模型）

每种模型动态响应的测定方法是：给模型外加一个阶跃变化的作用力 F，同时检测其所产生的长度变化（即位移 x）。在恒定的应力作用下，粘性材料的长度会缓慢变化，此过程称为蠕变（creep）。反过来，如果施加一个位移 x 的阶跃变化信号，然后检测其引起的应力 F 的变化，那么，在位移 x 保持恒定的状态下，粘性材料的应力 F 会逐渐下降，这个过程称为应力弛豫（stress relaxation），也称应力松弛。通过设计 Simulink 程序很容易计算各种动态模型的蠕变和应力弛豫。图 4.17 所示是没有粘性的纯弹性材料在阶跃作用力 F 作用下所产生的位移 x 的变化，显然这种材料没有蠕变。图 4.18 所示是纯弹性材料受到位移 x 的阶跃输入时所产生的作用力 F 的变化，同样，x 的变化立即导致 F 的阶跃响应，它没有应力弛豫过程。

James Clerk Maxwell（1831-1879）使用理想弹簧和理想阻尼器的组合描述了空气的粘弹性特性。图 4.19 所示就是 Maxwell 粘弹性模型在阶跃作用力的作用下所产生的蠕变响应，图 4.20 则显示了其应力弛豫响应。从图 4.19 可见，系统的蠕变响应

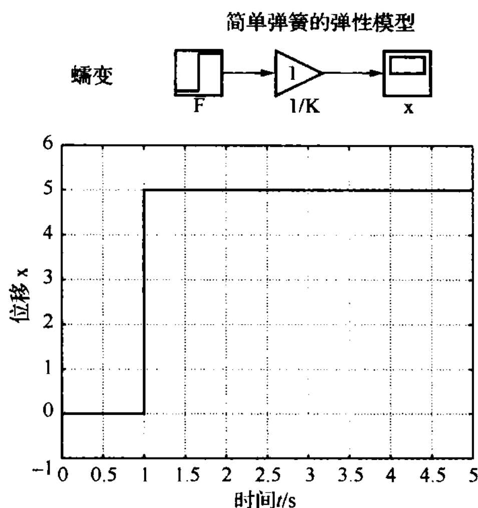


图 4.17　上图：纯弹性材料（即理想弹簧）的蠕变测试 Simulink 仿真模型及其输出结果（该模型的方程是 $x=F/K$，其中 x 为位移，F 为作用力，K 为弹簧系数）；下图：$K=1$ 且作用力从 0 阶跃到 5（任意单位）时材料的位移输出响应，x 与 F 呈线性关系

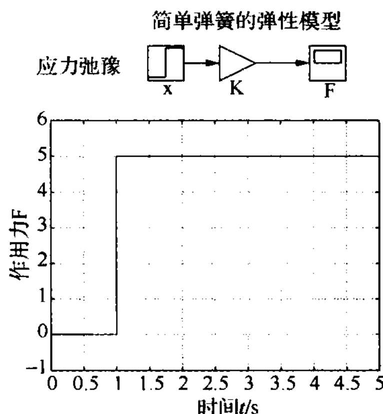


图 4.18　上图：纯弹性模型的应力弛豫测试 Simulink 仿真模型及其输出结果（该模型的方程是 $F=Kx$）；下图：弹簧系数 $K=1$ 且施加阶跃位移 $x=5$ 时的应力响应（其应力与位移呈线性关系）

几乎没有位移限制，理想阻尼器可以无限伸长。从图 4.20 可见，阶跃位移立即导致系统产生一个很大的初始应力，然后该应力逐渐弛豫，恢复到原有的应力水平。

Woldemar Voight（1850—1919）曾经在晶体学研究中使用理想弹簧与理想阻尼器的并联模型。图 4.21 所示是 Voight 粘弹性模型的蠕变测试结果，图 4.22 的应力弛豫响应则表明该模型的作用力可以无穷大，也就是位移发生阶跃变化时，由于阻尼器不能即刻产生响应，致使系统的作用力变为无穷大。

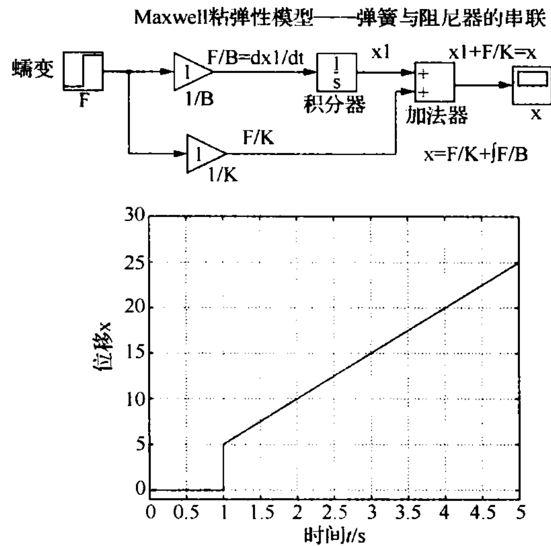


图 4.19　Maxwell 粘弹性模型的蠕变

注：该模型由理想弹簧和理想阻尼器串联组成（参见图 4.16）。设弹簧系数 $K=1$ 且阻尼器的阻尼系数 $B=1$（单位任意）。系统在阶跃变化力的作用下，位移 x 的响应是方程 $x=F/K+\int F/B$ 的解。x 响应曲线上首先是弹簧产生的即刻位移变化，紧接着是理想阻尼器的蠕变。只要作用力存在，这种蠕变一直可以延伸下去。

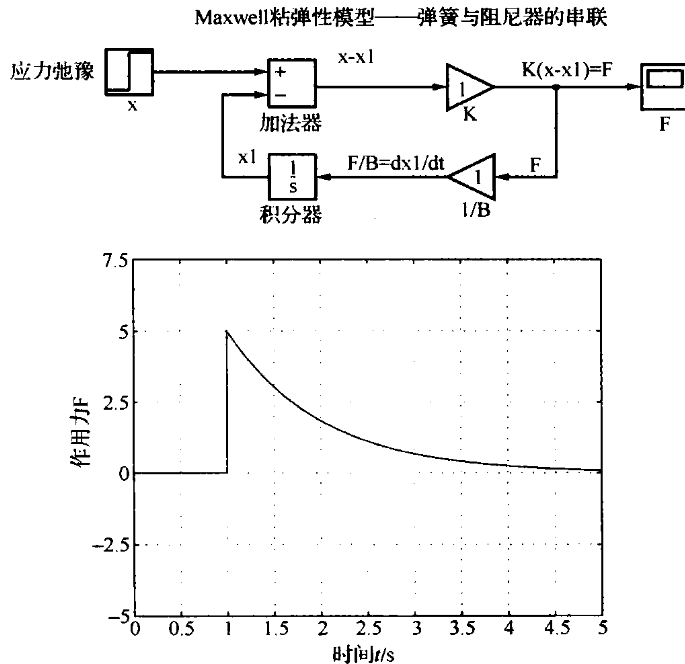


图 4.20　Maxwell 粘弹性模型的应力弛豫响应

注：该响应是方程 $F=K[x-\int F/B]$ 的解。设 $K=B=1$（单位任意），那么，阶跃位移导致的应力响应曲线上首先出现的是理想弹簧的即刻应力增量，紧接着是阻尼器的应力弛豫，使应力逐渐回复到稳态水平。

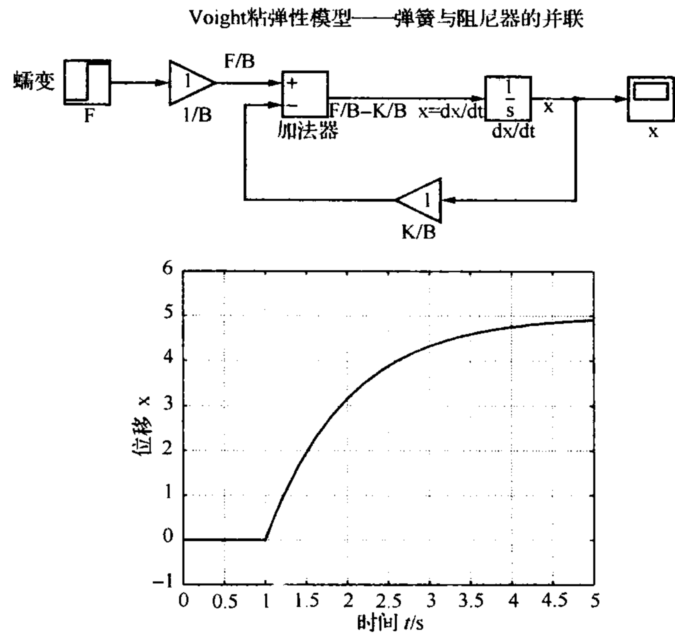


图 4.21 Voight 粘弹性模型的蠕变响应

注：该模型是理想弹簧与理想阻尼器的并联（参见图 4.16b)．其方程是 $\mathrm{d}x/\mathrm{d}t=1/B\ [F-Kx]$。当 $K=B=1$，且施加阶跃作用力 5（单位任意）时，其蠕变响应如图中曲线所示，位移逐渐蠕变到稳态值。

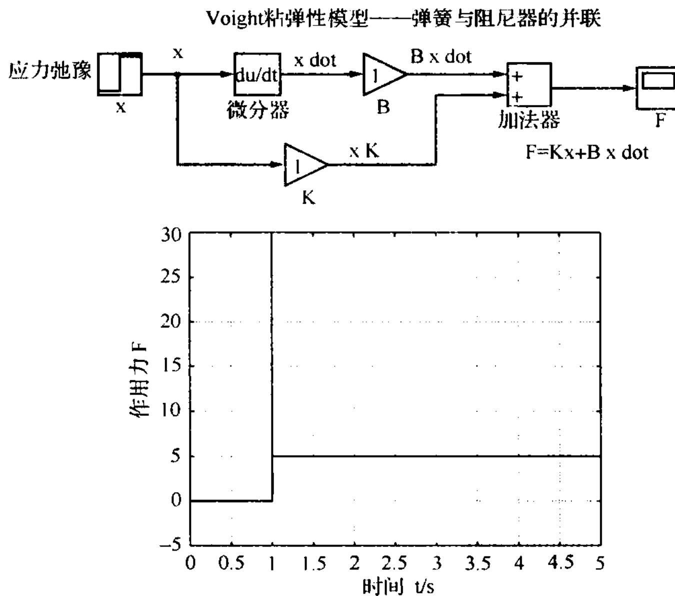


图 4.22 Voight 粘弹性模型的应力弛豫响应

注：该模型的方程是 $F=Kx+B\mathrm{d}x/\mathrm{d}t$。由于阻尼器与弹簧并联，并且它不能即刻响应位移的阶跃变化，致使模型的作用力瞬时变成无穷大。

William Thompson（Lord Kelvin　1824—1907）曾经用图 4.16c 所示的 3 个元件组成的 Kelvin 粘弹性模型，描述了各种不同固体材料的扭摆形式的力学特性。图 4.23 所示是 Kelvin 模型的蠕变响应，该响应有一个快速的初始跃变，紧接着是缓慢的蠕变。图 4.24 所示是 Kelvin 模型的应力弛豫响应，初始时，该响应表现出刚性变化，随后由于阻尼器的伸展，应力逐渐衰减到某个非 0 稳态水平。Kelvin 三元件模型是最简单的集中参数粘弹性模型，其位移延伸量和作用力都是有限值。

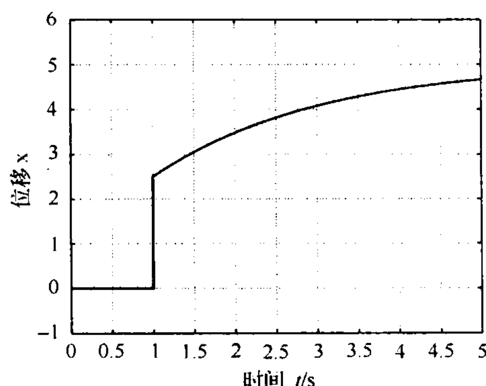


图 4.23　Kelvin 三元件粘弹性模型的蠕变响应

注：模型方程留给读者自己推导。作用力发生阶跃变化之后，模型产生一个瞬间的初始位移增量，接着逐渐蠕变到稳态水平。

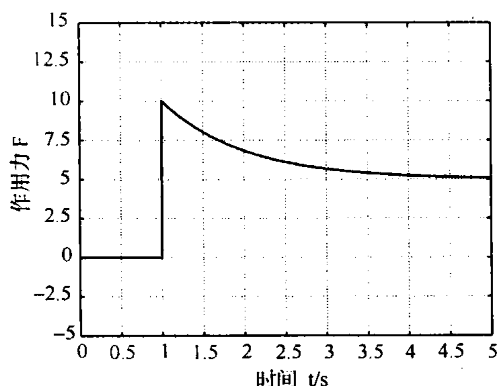


图 4.24　Kelvin 粘弹性模型的应力弛豫响应

注：模型首先产生一个瞬间的初始作用力增量，接着逐渐弛豫到稳态水平。

不过，这种三元件粘弹性模型只描述了应力弛豫和蠕变的基本特性，生物材料的粘弹性往往要复杂得多。例如，利用集中参数粘弹性模型只能计算不同频率外加应变作用下的离散的滞后曲线，而生物组织却表现出很宽的分布式滞后特性。这个问题的一种解决方法就是用三元件模型的分布式网络来描述生物材料。另一种方法是用 Westerhof 和 Noordergraaf（1990）提出的通用粘弹性模型来描述血管壁等生物材料的粘弹性特性。也就是根据动脉血管的特性，将弹性模量设计成随频率变化的复杂参量，使模型包含频率依赖性的弹性模量、应力弛豫、蠕变和滞后特性。这样，Maxwell 模型和 Voight 模型就只是这种通用模型的特例。

4.5　软骨、韧带、肌腱与肌肉

骨骼关节的表面覆盖着软骨，软骨主要由胶原蛋白组成，胶原蛋白是构成动物体内各种软硬组织的主要材料。分离出来的胶原纤维具有很大的抗拉强度，与尼龙的抗拉强度（50～100MPa）相当，且弹性模量接近 1GPa。还有一种在脊椎动物体内发现的弹性蛋白，它在血管壁和肺中具有特别重要的作用，是所有已知固体生物材料中弹性最好的材料，其弹性模量约为 0.6MPa。弹性蛋白使皮肤和结缔组织具有很好的弹性。

4.5.1 软骨

软骨构成了关节的承重表面，它是一种多孔渗水的粘弹性材料。在关节承受负荷的过程中，在应力作用下，体液会在软骨组织中流进流出，从而使软骨表现出复杂的力学特性。这种特性使关节软骨的粘弹性非常强，其受压应力弛豫时间有1～5s数量级。软骨是各向异性材料，在加载和卸载的循环周期中存在滞后现象。其极限压应力约在5MPa数量级。

4.5.2 韧带与肌腱

韧带将骨骼连接在一起，从而构成生物体的骨架。肌腱则将肌肉连接到骨骼上，并将肌肉收缩产生的力传给骨骼，从而带动肢体的运动。韧带和肌腱的主要作用是传递拉力，它们都由平行的胶原纤维束组成，并具有相似的力学特性。人体肌腱的极限应力约为50～100MPa，并具有明显的非线性应力－应变关系曲线。曲线中线性区域内的弹性模量约为1～2GPa。韧带和肌腱都具有滞后、粘弹性蠕变以及应力弛豫性质。另外，这两种材料的响应都与初始状态相关，因此，通过预先加载作用力的“预处理”方法可以影响随后的应力－应变响应关系。经过反复加载的处理，这两种生物组织内部的结构会发生变化，从而导致材料特性的偏移。

4.5.3 肌肉力学

本书第3章已经介绍过，肌肉是一种主动的可兴奋组织，它通过相互交叉的肌动蛋白丝与肌球蛋白丝之间形成的横桥连接来产生作用力。描述肌肉收缩的定量模型有两种：一是A. V. Hill建立的集中参数收缩模型；二是A. F. Huxley建立的描述单个肌小节的横桥模型（参见Palladino和Noordergraaf 1998年的著作）。最早的肌肉模型只是用单纯的弹簧简单描述整块肌肉的力学特性，弹簧受到拉伸时所储存的势能在缩短时就释放出来。Ernst Weber（1846）最先提出肌肉具有顺应性，他认为肌肉是一种弹性材料，在其兴奋过程中，通过化学能的转化，肌肉的状态会发生变化。此后，研究人员都只关注肌肉的弹性特性，却忽视了新陈代谢引起的肌肉刚性的变化。单纯的弹性模型受到质疑，因为它不符合热力学定律。肌肉弹簧拉伸时所储存的势能小于它缩短时所释放的能量，这个释放的能量包括弹簧的做功和发热。于是，人们用传统弹簧与阻尼器的组合来描述肌肉。1922年Hill将弹簧与粘弹性元件组合在一起，这才在肌肉模型中引入了早在1940年代就曾经提出的粘弹性。

快速拉伸和释放实验表明，肌肉的粘弹性与时间具有很强的相关性。一般，肌肉长度变化发生得越快，其收缩力受到的影响就越大。显然，带阻尼的弹簧系统并不能完整描述肌肉收缩的复杂机制。1935年，Fenn和Marsh在Hill的阻尼弹性模型中添加了一个串联弹性元件，并指出“简单的力学系统不能正确模拟肌肉的特性”。后来，Hill直接用实验获得的双曲线表示负荷与骨骼肌初始收缩速度之间的关系，并将该双曲线作为模型的一个模块，代表一种收缩元件。Hill原来的粘弹性模型认

为肌肉具有恒定的势能，该势能的释放速率由粘弹性决定。新模型不用摩擦机制，而用一种未明确定义的内部机制控制能量的释放。这种引入随负荷变化的肌肉动力学新特性是一个正确的研究方向；但是，后来的模型（包括用于产热研究的模型）所使用的双曲线都基于骨骼肌强直收缩实验的数据。这种方法有两个缺点：（1）这种单一的力一速度关系曲线使得所有具体应用模型中的收缩元件都具有相同的肌肉能量和作用力特性；（2）这种建模方法不能揭示有关肌肉收缩机制的信息。由于缺乏能够描述特定负荷状态的收缩元件，研究人员在模型中不断增加被动弹簧和阻尼器的数目。到了 1960 年代末，肌肉模型中的元件数目至少已经增加到了 9 个。至今，肌肉模型设计中都使用基于 Hill 收缩元件的分布式模型，最新的模型主要用于描述肌肉的强直收缩、等长收缩以及等长痉挛收缩。

另一类肌肉收缩模型的研究工作在超微结构水平进行，这些研究以肌丝滑行理论为基础。肌丝滑行是公认的肌肉收缩机制，肌肉产生作用力的过程是粗肌丝与细肌丝之间通过横桥的结合将化学能转化为机械能的过程。关于横桥结合和分离的详细模式至今仍然存在争议，特别是肌肉松弛的机制尚不清楚。早在实际观察到横桥之前，1957 年 A. F. Huxley 就已经根据有关肌肉结构和能量的假设建立了横桥的模型，他用速率常数 f 和 g 分别描述粗细肌丝之间结合和分离两个过程。这个模型的主要缺点是不能解释肌肉长度或者负荷快速变化所引起的瞬间突变，就像前面有关蠕动和应力弛豫的 Simulink 仿真结果中所显示的即刻变化。

后来出现了一些改进模型，它们用复杂的速率函数描述横桥的结合和分离过程，这些模型往往只适用于描述单对肌丝。现在，各种肌肉模型都趋向于只描述单纯的一种实验现象，如快速释放实验等。没有一种模型能够广泛适用于各种肌肉收缩情况。描述横桥结合和分离的速率函数则越来越复杂。原来建立肌肉模型是利用简单的横桥机制来描述复杂的肌肉动力学；而现在的趋势却是反过来，变成利用复杂的横桥动力学特性来描述特定的实验结果。

另一方面，Palladino 和 Noordergraaf 在 1998 年提出了一种大规模的分布式肌肉模型，该模型基于横桥连接的基本力学特性来模拟肌肉的收缩和松弛过程，其复杂的肌肉收缩特性来自于肌肉超微结构的动力学特性，也就是结构决定功能。肌丝之间的结合用粘弹性材料来模拟。肌肉收缩用电刺激诱发，电信号在心肌中以一定的速度传播，在肌肉空间中形成了非同步刺激。另外，肌细胞内质网中释放的钙离子需要经过扩散才能到达肌球蛋白头部。这些过程以及肌肉结构的非均匀性都有力地表明收缩在整块肌肉中是非同步的。这种模型抛弃了肌肉收缩的同步假设，使用分布式特性，并考虑肌丝的质量，这样就可以使粗肌丝相对细肌丝产生很小的移动。这种移动导致横桥结合的分离并产生热量。肌肉的总体收缩由此产生。该模型可以模拟实验中观察到的外加负荷致使肌肉长度发生快速突变的现象。肌肉的松弛则可以由模型的分布特性产生。

这种分布式的肌肉新模型具有以下主要特点：肌小节由相互重叠的粗肌丝和细肌丝构成，粗、细肌丝之间通过横桥连接，肌细胞兴奋时横桥将两种肌丝结合在一

起，肌细胞松弛时横桥分离，两种肌丝不再相连。图 4.25 显示了由一段段肌小节连接而成的肌细胞模型。其中的每个横桥用三元件固体粘弹性模型描述，肌丝则表示成质量。肌肉产生的作用力由相互交错的肌丝之间的粘弹性横桥连接形成。横桥数目由粗、细肌丝之间重叠的程度决定，它是时间和空间的函数，与电激活和化学激活的速率相关。横桥形成的非同步性、各个肌小节中横桥数目的差别以及肌肉的收缩和外加的长度突变等，都会引起肌丝的微小移动。由于肌丝存在质量，这种移动具有阻尼振荡形式，其振荡频率由分布式系统的特性决定。当横桥结合的作用力降为 0 时，肌丝间的横桥分离。因此，肌丝的运动和横桥应力的弛豫导致了横桥的分离并导致肌肉松弛，这里没有用到横桥分离的速率函数。事实上，肌肉松弛由系统固有的不稳定性产生。虽然这种模型由线性非时变元件（即弹簧、阻尼器和质量）组成，但是，模型的动力学结构却使其力学特性表现出高度的非线性和时变性，这与实际肌纤维的情况一致。

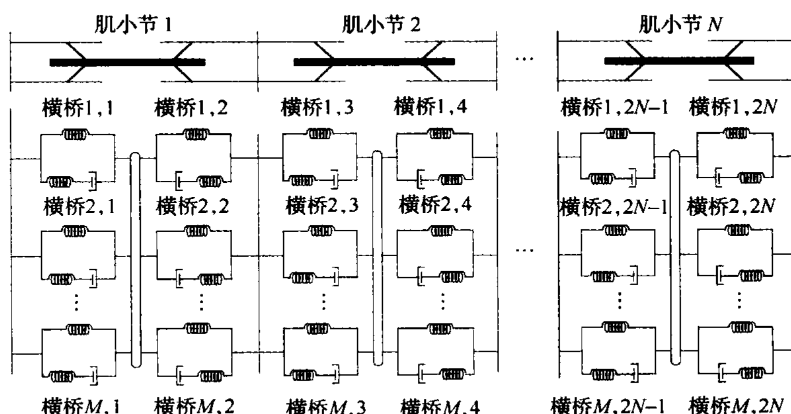


图 4.25 由 N 个肌小节的分布式网络构成的肌细胞模型

注：其中每个肌小节包含 M 对并列的横桥连接。（摘自 Palladino 和 Noordergraaf 1998 年著作第 44 页的图 3.4）

该模型对于机械扰动的灵敏度与实验结果一致。实验的肌张力记录曲线、用水母素发光蛋白测定肌细胞内游离钙离子浓度以及高速 X 射线衍射研究都表明了横桥分离的作用。该模型与肌小节长度的反馈研究结果也一致，并可以在实验前预测肌细胞的动力学过程。肌小节长度的反馈可以通过减小内部移动来延缓肌肉的松弛。

该模型提出了肌肉复杂力学特性的结构机制，并且预测了收缩机制的新特点，如肌肉松弛的机制、肌肉发热的估算等。这种新模型将肌肉的复杂力学特性建立在其物理解剖结构上，因此，它将亚细胞结构与组织层次的功能联系起来了。

本章只介绍生物组织力学特性的一些要点，更全面的内容请参阅冯元桢教授的《生物力学：活组织的力学特性》(Biomechanics: Mechanical Properties of Living Tissues)、Nigg 与 Herzog 合著的《肌肉骨骼系统的生物力学》（Biomechanics of the Musculo—Skeletal System）以及 Mow 与 Hayes 合著的《整形外科生物力学基础》

(Basic Orthopaedic Biomechanics)等著作。肌肉收缩研究的悠久历史在 Needham 的著作《肉机器》(Machina Carnis)中有详细的论述。有关的医学历史请参阅 Singer 与 Underwood 在 1962 年合著的《医学简史》(A Short History of Medicine)。本章下面两节将介绍以上 4.2～4.5 节所述的生物力学原理在人体步态分析和心血管系统定量研究中的应用。

4.6 临床步态分析

动力学在人体运动分析中的应用之一就是临床步态分析。临床步态分析包括病人步行模式特征参数的测定、数据的解释以及治疗方案的制定。这是一项高度合作性的工作，需要病人与多个领域专家之间的配合，尤其是医生、理疗师、人体运动学专家和工程师(或者技师)。其中，工程师的任务很艰巨，他们要根据现有的仪器条件建立临床步态分析的检测系统。数据测量和采集的根本目标是在不改变病人自然活动的状态下，准确检测病人的运动，使数据具有足够的精度，可以用于临床分析。虽然临床步态测量的部分仪器设备已有商品化的产品，但是，设备的使用方案至今仍一直在不断改进，测量方案及其相关模型的有效性最终决定了检测数据的意义和质量。可见，工程人员与相关临床伙伴之间的合作会对临床步态分析的过程产生重要的影响。

通常，采集步态数据时，要在病人皮肤表面放置高反光标志物，然后，利用安装在周围环境中的一组视频运动摄像机拍摄这些体表反光标志物。根据两台以上摄像机同时拍摄的图像，可以计算出每个时刻下这些标志物的空间位置。除了空间位置信息以外，步态分析实验中还可以检测其他参数。例如，利用地面下安装的测力平台可以测量地面受到的作用力；用体表电极或者细导线肌电极可以检测肌电图(所用的电极类型由具体的被测肌肉决定)。

为了与本章所述主题一致，我们下面主要介绍临床步态分析的生物力学部分，包括人体各个肢体段和关节的运动学分析和关节动力学分析，以及临床数据含义的简单解释。

4.6.1 临床步态模型

步态模型是指将步行试验中采集的数据转换成临床诊断信息的算法。例如，通过步态模型可以根据病人身上标志点的三维空间位置数据计算出各种角度参数，用于描述病人各肢体段及其末端关节的运动过程。只有充分了解临床诊断的需求，才能设计出合适的步态模型。这里，步态模型的设计既要考虑测量设备的技术条件限制，也要考虑不同的试验对象，以便满足不同病人群体的需要，如具有不同年龄、不同步态疾病、不同步行能力的病人群体。好的模型必须能够广泛适用于各种不同病人，例如，具有不同体格和智力的成人和儿童；并且要有足够的精确度，可以解决生物力学的细节问题；还要可重复，具备临床实用性。

4.6.2 运动学分析

步态测试时，测试系统的视频摄像机记录下粘贴在病人皮肤表面的反光标志点的空间和时间运动轨迹，然后这些标志点的轨迹再用于各个坐标系的计算。这些坐标系与所需分析的人体肢体段的解剖位置相对应，是计算人体各个肢体段绝对空间方位以及段与段之间角位移（如关节角等）的基础。为了完成这些计算，人体的每个肢体段至少需要用 3 个不在同一直线上的参考标志点来标定，这些标志点形成一个平面，用于推导各个肢体段的坐标系。理论上，通过 3 个标志点的方法可以将人体任意肢体段的运动过程记录下来；但是，实际上只有当这些标志点与人体的解剖结构具有密切联系时，这些运动学数据才具有临床应用价值。标志点要么直接贴在肢体段上骨骼特征明显的部位，要么放在便于摄像机拍摄同时又与体内骨骼具有明确对应关系的部位。下面用人体骨盆和大腿这两个肢体段分别说明这两种情况。

骨盆的解剖坐标系

采集骨盆的运动学数据时，其中两个标志物分别贴在左右两侧髂前上棘（Anterior Superior Iliac Spine，ASIS）处，还有一个标志物可以贴在左或右髂后上棘（Posterior Superior Iliac Spine，PSIS）处，这样，就可以用拍摄到的标志物的空间位置数据计算骨盆的解剖坐标系。详见以下例题。

例 4.9 以下是一组骨盆标志点的三维位置数据，单位为 m，坐标系是图 4.26

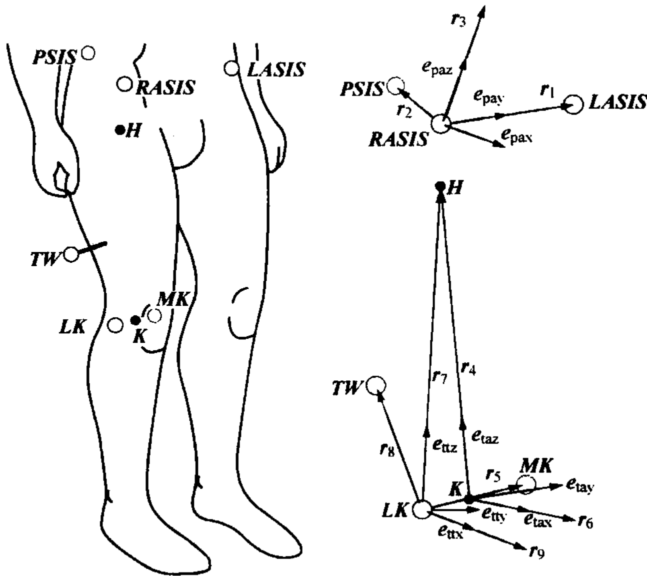


图 4.26 用于定义骨盆和大腿坐标系的标志点

注：骨盆的标志点中，***PSIS*** 表示髂后上棘，***H*** 表示髋关节旋转中心，***LASIS*** 和 ***RASIS*** 分别表示左、右髂前上棘。大腿的标志点中，***TW*** 表示大腿杆，***K*** 表示膝关节中心，***MK*** 和 ***LK*** 分别为膝关节（即股骨髁）的内、外侧标志点。

所示的实验室空间固定坐标。

$$\text{右 ASIS：} \boldsymbol{RASIS} = -0.850\boldsymbol{i} - 0.802\boldsymbol{j} + 0.652\boldsymbol{k}$$

$$\text{左 ASIS：} \boldsymbol{LASIS} = -0.831\boldsymbol{i} - 0.651\boldsymbol{j} + 0.652\boldsymbol{k}$$

$$\boldsymbol{PSIS} = -1.015\boldsymbol{i} - 0.704\boldsymbol{j} + 0.686\boldsymbol{k}$$

请计算骨盆的解剖坐标系。

解：题中给出的 3 个标志点构成了一个平面，左右 ASIS 之间的连线是骨盆坐标系中的一个轴，另一个坐标轴垂直于骨盆平面，第三个坐标轴则与前两个坐标轴正交。因此，骨盆解剖坐标系三个坐标轴的求解步骤如下：

（1）首先计算第一个坐标轴。将矢量 $\boldsymbol{LASIS}$ 减去矢量 $\boldsymbol{RASIS}$：

$$\boldsymbol{LASIS} - \boldsymbol{RASIS} = (-0.831 - (-0.850))\boldsymbol{i} + (-0.651 - (-0.802))\boldsymbol{j} + (0.652 - 0.652)\boldsymbol{k}$$

可得矢量

$$\boldsymbol{r}_1 = 0.0190\boldsymbol{i} + 0.1510\boldsymbol{j} + 0.0000\boldsymbol{k}$$

其单位矢量就是

$$\boldsymbol{e}_{r1} = \frac{0.019\boldsymbol{i} + 0.151\boldsymbol{j} + 0.000\boldsymbol{k}}{\sqrt{0.019^2 + 0.151^2 + 0.000^2}}$$

$$\boldsymbol{e}_{r1} = 0.125\boldsymbol{i} + 0.992\boldsymbol{j} + 0.000\boldsymbol{k}$$

如图 4.26 所示，该单位矢量 $\boldsymbol{e}_{r1}$ 就表示骨盆解剖坐标系的内外侧方向 $\boldsymbol{e}_{pay}$（即 y 轴）。

（2）然后计算第二个坐标轴，即垂直于骨盆平面的矢量。将 $PSIS$ 减去 $RASIS$，可得

$$\boldsymbol{r}_2 = -0.165\boldsymbol{i} + 0.098\boldsymbol{j} + 0.034\boldsymbol{k}$$

求矢量叉积 $\boldsymbol{e}_{pay} \times \boldsymbol{r}_2$，可得

$$\begin{aligned}
\boldsymbol{r}_3 &= \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 0.125 & 0.992 & 0.000 \\ -0.165 & 0.098 & 0.034 \end{vmatrix} \\
&= [(0.992)(0.034) - (0.000)(0.098)]\boldsymbol{i} + \\
&\quad [(0.000)(-0.165) - (0.125)(0.034)]\boldsymbol{j} + \\
&\quad [(0.125)(0.098) - (0.992)(-0.165)]\boldsymbol{k} \\
&= 0.034\boldsymbol{i} - 0.004\boldsymbol{j} + 0.176\boldsymbol{k}
\end{aligned}$$

其单位矢量为

$$\boldsymbol{e}_{r3} = \boldsymbol{e}_{paz} = 0.188\boldsymbol{i} - 0.024\boldsymbol{j} + 0.982\boldsymbol{k}$$

该单位矢量 $\boldsymbol{e}_{r3}$ 就表示骨盆解剖坐标系的前上方向 $\boldsymbol{e}_{paz}$（即 z 轴）。

（3）最后求第三个坐标轴（即 x 轴），该坐标轴与前两个轴正交。求矢量叉积 $\boldsymbol{e}_{pay} \times \boldsymbol{e}_{paz}$ 可得 x 轴的方向为

$$\boldsymbol{e}_{pax} = 0.974\boldsymbol{i} - 0.123\boldsymbol{j} - 0.190\boldsymbol{k}$$

于是，本例的骨盆解剖坐标系可以表示为

$$\{e_{pa}\}=\begin{bmatrix}e_{pax}\\e_{pay}\\e_{paz}\end{bmatrix}=\begin{bmatrix}0.974 & -0.123 & -0.190\\0.125 & 0.992 & 0.000\\0.188 & -0.024 & 0.982\end{bmatrix}\begin{bmatrix}i\\j\\k\end{bmatrix}$$

注意，这3个坐标轴的系数就是定义骨盆坐标系与实验室坐标系之间关系的方向余弦。■

总之，通过监测3个骨盆标志点的运动轨迹，就可以求得每个时刻下由 e_{pax}、e_{pay} 和 e_{paz}3个坐标轴构成的骨盆坐标系 $\{e_{pa}\}$ 的方位。再运用式（4.32）～式（4.34）的欧拉角计算公式，就可以求出该坐标系的绝对角位移，并用于分析骨盆的倾斜和旋转。本节后面将举例说明这种计算角位移的方法。

大腿的解剖坐标系

要在大腿上找出用作步态检测参考点的3个骨骼特征点，比骨盆要困难。一种看似可行的方法是将标志物贴在股骨髁的内外两侧和大转子处，但是，走动时贴在股骨髁内侧的标志并不一定可行，有些病人行走时两个膝盖之间会摩擦。另外，文献中经常提到在大转子处粘贴标志物；其实，由于走动时皮肤会产生明显的移动，外部标志物与皮肤下大转子之间的相对位移很大，因此，实际上也不能用作参考点。

一般，定量分析大腿、小腿和脚的活动时，经常在人体某个静止姿势下在肢体段上增加一些解剖位置的标志点，这些标志点称为静态解剖标志点，在走动测试时要去掉，而那些走动时仍带着的标志物则称为运动标志点。并且，这两类标志点之间具有确定的数学对应关系，这些对应关系在步态测试过程中保持不变，也就是假设人体各个肢体段是刚性的。下面举例说明这种方法。

例4.10 假设病人静止站立时各个标志点的坐标数据如下，单位为m：

股骨髁外侧标志点 $\boldsymbol{LK}=-0.881\boldsymbol{i}-0.858\boldsymbol{j}+0.325\boldsymbol{k}$

股骨髁内侧标志点 $\boldsymbol{MK}=-0.855\boldsymbol{i}-0.767\boldsymbol{j}+0.318\boldsymbol{k}$

请计算大腿的解剖坐标系。

解：大腿平面由以下3个解剖标志点构成：髋关节旋转中心和股骨髁内、外侧标志点 $\boldsymbol{MK}$ 和 $\boldsymbol{LK}$。膝关节中心位置 $\boldsymbol{K}$ 可以用 $\boldsymbol{LK}$ 与 $\boldsymbol{MK}$ 之间的中点来表示。膝关节中心与髋关节旋转中心这两点形成的矢量就是大腿坐标系的纵轴。第二个坐标轴垂直于大腿平面，第三个坐标轴则与前两个轴正交。

首先计算膝关节中心位置：

$$\frac{\boldsymbol{LK}+\boldsymbol{MK}}{2}=\frac{(-0.881)+(-0.855)}{2}\boldsymbol{i}+\frac{(-0.858)+(-0.767)}{2}\boldsymbol{j}+\frac{(0.325)+(0.318)}{2}\boldsymbol{k}$$

可得

$$\text{膝关节中心位置 } \boldsymbol{K}=-0.868\boldsymbol{i}-0.812\boldsymbol{j}+0.321\boldsymbol{k}$$

股骨头的中点就是髋关节旋转中心 $\boldsymbol{H}$。根据人体测量学的骨盆结构统计模型，可以得到 $\boldsymbol{H}$ 位置的近似数据（Davis，1991），即：

$$\text{髋关节旋转中心的位置 } \boldsymbol{H}=-0.906\boldsymbol{i}-0.763\boldsymbol{j}+0.593\boldsymbol{k}$$

详细内容这里不作介绍。

现在，就可以按照如下步骤计算大腿的解剖坐标系：

1）$\boldsymbol{H}$ 矢量减 $\boldsymbol{K}$ 矢量，可得

$$\boldsymbol{r}_4=-0.038\boldsymbol{i}+0.049\boldsymbol{j}+0.272\boldsymbol{k}$$

其单位矢量为

$$\boldsymbol{e}_{r4}=\boldsymbol{e}_{taz}=-0.137\boldsymbol{i}+0.175\boldsymbol{j}+0.975\boldsymbol{k}$$

该单位矢量表示坐标系的纵轴方向 $\boldsymbol{e}_{taz}$，即 z 轴。

2）与骨盆坐标系相似，大腿平面的第二个矢量是垂直于该平面的矢量。因此，$\boldsymbol{MK}$ 矢量减 $\boldsymbol{LK}$ 矢量，可得

$$\boldsymbol{r}_5=0.026\boldsymbol{i}+0.091\boldsymbol{j}-0.007\boldsymbol{k}$$

求矢量叉积 $\boldsymbol{r}_5\times\boldsymbol{e}_{taz}$ 可得

$$\boldsymbol{r}_6=0.090\boldsymbol{i}-0.024\boldsymbol{j}+0.017\boldsymbol{k}$$

其单位矢量为

$$\boldsymbol{e}_{r6}=\boldsymbol{e}_{tax}=0.949\boldsymbol{i}-0.258\boldsymbol{j}+0.180\boldsymbol{k}$$

该单位矢量表示坐标系的前后方向 $\boldsymbol{e}_{tax}$，即 x 轴。

3）再计算与前两个坐标轴正交的第三个坐标轴，即由叉积 $\boldsymbol{e}_{taz}\times\boldsymbol{e}_{tax}$，求坐标系的 y 轴矢量 $\boldsymbol{e}_{tay}$：

$$\boldsymbol{e}_{tay}=0.284\boldsymbol{i}+0.950\boldsymbol{j}-0.131\boldsymbol{k}$$

于是，大腿的解剖坐标系可以表示为：

$$\{\boldsymbol{e}_{ta}\}=\begin{bmatrix}\boldsymbol{e}_{tax}\\\boldsymbol{e}_{tay}\\\boldsymbol{e}_{taz}\end{bmatrix}=\begin{bmatrix}0.949 & -0.258 & 0.180\\0.284 & 0.950 & -0.131\\-0.137 & 0.175 & 0.975\end{bmatrix}\begin{bmatrix}\boldsymbol{i}\\\boldsymbol{j}\\\boldsymbol{k}\end{bmatrix}$$

它由 $\boldsymbol{e}_{tax}$、$\boldsymbol{e}_{tay}$ 和 $\boldsymbol{e}_{taz}$ 3 个坐标轴组成。在此坐标系的建立过程中使用了股骨髁内侧的标志点，该标志在行走测试时要去掉。因此，膝关节中心的位置不能按照上述方法计算，解决这个难题的方法是：在大腿表面放置另外一个标志点，使其与髋关节旋转中心和股骨髁外侧标志点一起组成一个平面，这 3 个参考点就可以用于计算大腿的一个“技术”坐标系，再由这个坐标系计算膝关节中心的位置。具体方法见以下例题。■

例 4.11　继续以上例 4.10，给定大腿上另一个标志点的坐标（它不对应于骨骼的特定解剖结构）：

大腿标志点 $\boldsymbol{TW}=-0.890\boldsymbol{i}-0.937\boldsymbol{j}+0.478\boldsymbol{k}$

请计算大腿的“技术”坐标系。

解：大腿“技术”坐标系的计算步骤如下：

1）计算坐标系的纵轴方向矢量 $\boldsymbol{e}_{tt}$，即 z 轴，先将髋关节旋转中心 $\boldsymbol{H}$ 矢量减 $\boldsymbol{LK}$ 矢量，可得

$$\boldsymbol{r}_7=-0.025\boldsymbol{i}+0.094\boldsymbol{j}+0.268\boldsymbol{k}$$

其单位矢量为

$$\boldsymbol{e}_{r7}=\boldsymbol{e}_{ttz}=-0.088\boldsymbol{i}+0.330\boldsymbol{j}+0.940\boldsymbol{k}$$

该单位矢量表示坐标系的 z 轴 $\boldsymbol{e}_{ttz}$。

2）计算与 **LK**、**H** 和 **TW** 3 点组成的平面相垂直的坐标轴，即 **TW** 矢量减 **LK** 矢量，可得

$$\boldsymbol{r}_8=-0.009\boldsymbol{i}-0.079\boldsymbol{j}+0.153\boldsymbol{k}$$

再计算矢量叉积 $\boldsymbol{r}_7\times\boldsymbol{r}_8$，可得

$$\boldsymbol{r}_9=0.036\boldsymbol{i}+0.001\boldsymbol{j}+0.003\boldsymbol{k}$$

其单位矢量为

$$\boldsymbol{e}_{r9}=\boldsymbol{e}_{ttx}=0.996\boldsymbol{i}+0.040\boldsymbol{j}+0.079\boldsymbol{k}$$

该单位矢量就表示大腿坐标系的前后方向 $\boldsymbol{e}_{ttx}$，即 x 轴。

3）计算与前两个坐标轴正交的第三个坐标轴，即由叉积 $\boldsymbol{e}_{ttz}\times\boldsymbol{e}_{ttx}$ 求坐标系的 y 轴矢量 $\boldsymbol{e}_{tty}$：

$$\boldsymbol{e}_{tty}=-0.012\boldsymbol{i}+0.943\boldsymbol{j}+0.333\boldsymbol{k}$$

于是，大腿的技术坐标系可以表示为

$$\{\boldsymbol{e}_{tt}\}=\begin{bmatrix}\boldsymbol{e}_{ttx}\\\boldsymbol{e}_{tty}\\\boldsymbol{e}_{ttz}\end{bmatrix}=\begin{bmatrix}0.996 & 0.040 & 0.079\\-0.012 & 0.943 & 0.333\\-0.088 & 0.330 & 0.940\end{bmatrix}\begin{bmatrix}\boldsymbol{i}\\\boldsymbol{j}\\\boldsymbol{k}\end{bmatrix}$$ ■

注意，这个在站立校正状态下计算得到的坐标系 $\{\boldsymbol{e}_{tt}\}$，在每次行走测试之后也可以求得。计算该坐标系用到的股骨髁外侧、大腿杆部和髋关节旋转中心这 3 个标志点在站立和行走时都可以获得。然后，根据坐标系 $\{\boldsymbol{e}_{tt}\}$ 就可以求得其他实际的解剖坐标系。利用 4.2.2 节所介绍的方向余弦或者欧拉角都可以建立例 4.10 所述的大腿解剖坐标系 $\{\boldsymbol{e}_{ta}\}$ 与大腿技术坐标系 $\{\boldsymbol{e}_{tt}\}$ 之间的关系方程。并且，站立校正之后必须除去的那些标志点（如股骨髁内侧标志点 **MK**）以及膝关节中心等的解剖位置都可以先转换到技术坐标系下，然后，在分析步态数据时再还原回来。

肢体段与关节夹角

跟踪记录人体各个肢体段的解剖坐标系，就可以计算各肢体段的空间绝对方位角或者某个肢体段相对于另一个肢体段的夹角。如上述例 4.9 所示，利用 4.2.2 节所介绍的欧拉角式（4.32）～式（4.34），可以计算表示骨盆解剖坐标系 $\{\boldsymbol{e}_{pa}\}$ 与实验室坐标系（即固定坐标系）之间关系的 3 个欧拉角。注意，在以下计算公式中，无撇号的矢量表示实验室坐标系，带 3 个撇号的矢量表示骨盆解剖坐标系 $\{\boldsymbol{e}_{pa}\}$。由式（4.32）可得

$$\theta_x=-\arcsin(\boldsymbol{k}'''\cdot\boldsymbol{j})$$

即

$$\begin{aligned}\theta_x&=-\arcsin(\boldsymbol{e}_{paz}\cdot\boldsymbol{j})\\&=-\arcsin((0.188\boldsymbol{i}-0.024\boldsymbol{j}+0.982\boldsymbol{k})\cdot\boldsymbol{j})\\&=-\arcsin(-0.024)\end{aligned}$$

$$=1°\text{骨盆倾斜角}$$

同理，由式（4.33）可得

$$\theta_y=\arcsin\left(\frac{(\boldsymbol{k}''' \cdot \boldsymbol{i})}{\cos\theta_x}\right)$$

即

$$\theta_y=\arcsin\left(\frac{(\boldsymbol{e}_{paz} \cdot \boldsymbol{i})}{\cos\theta_x}\right)$$

$$=\arcsin\left(\frac{(0.188\boldsymbol{i}-0.024\boldsymbol{j}+0.982\boldsymbol{k})\cdot\boldsymbol{i}}{\cos 1°}\right)$$

$$=\arcsin\left(\frac{0.188}{\cos 1°}\right)$$

$$=11°\text{骨盆前倾角}$$

由式（4.34）可得

$$\theta_z=\arcsin\left(\frac{(\boldsymbol{i}''' \cdot \boldsymbol{j})}{\cos\theta_x}\right)$$

即

$$\theta_z=\arcsin\left(\frac{(\boldsymbol{e}_{pax} \cdot \boldsymbol{j})}{\cos\theta_x}\right)$$

$$=\arcsin\left(\frac{(0.974\boldsymbol{i}-0.123\boldsymbol{j}-0.190\boldsymbol{k})\cdot\boldsymbol{j}}{\cos 1°}\right)$$

$$=\arcsin\left(\frac{-0.123}{\cos 1°}\right)$$

$$=-7°\text{骨盆旋转角}$$

同样，用这些欧拉公式也可以计算表示大腿解剖坐标系 $\{\boldsymbol{e}_{ta}\}$ 与骨盆解剖坐标系 $\{\boldsymbol{e}_{pa}\}$ 之间关系的三个欧拉角。将例 4.9 和例 4.10 求得的 $\{\boldsymbol{e}_{pa}\}$ 和 $\{\boldsymbol{e}_{ta}\}$ 代入式（4.32）可得

$$\theta_x=-\arcsin(\boldsymbol{e}_{taz}\cdot\boldsymbol{e}_{pay})$$

$$=-\arcsin((-0.137\boldsymbol{i}+0.175\boldsymbol{j}+0.975\boldsymbol{k})\cdot(0.125\boldsymbol{i}+0.992\boldsymbol{j}+0.000\boldsymbol{k}))$$

$$=-\arcsin(0.156)$$

$$=-9°\text{髋关节外展或内收角}$$

负号是指左大腿的髋关节内收，或者指右大腿的髋关节外展。

另外，将 $\{\boldsymbol{e}_{pa}\}$ 和 $\{\boldsymbol{e}_{ta}\}$ 的值代入式（4.33）和式（4.34）可得

髋关节前屈或后伸　$\theta_y=20°$

髋关节内旋或外旋　$\theta_z=-8°$

对于髋关节内旋或外旋，负号是指左大腿的髋关节内旋，或者指右大腿的髋关节外旋。对于髋关节前屈或后伸，负号表示后伸，左右大腿一样。人体其他肢体段，如胫骨（即小腿）、脚、躯干、手臂和头等，只要有正确定义的解剖坐标系，都可以用这种处理方法。

4.6.3 动力学分析

如上所述，人体活动时标志点的位移轨迹（即运动数据）可以用于肢体段和关节的运动学分析。这些运动学数据还可以与地面反作用力和力矩等数据相结合，力和力矩的作用点称为压力中心点。再利用肢体段的重量和转动惯量的估计值，就可以计算关节受到的净作用力和力矩。下面详细介绍计算过程。

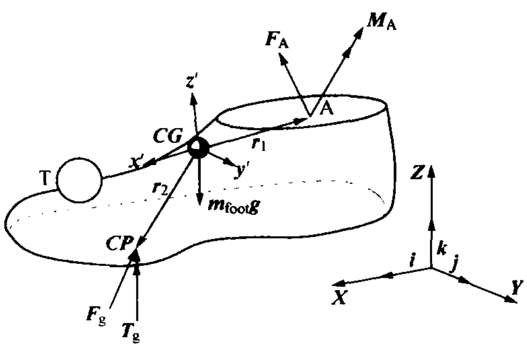


图 4.27 利用踝关节 A 和脚尖 T 的标志点数据，并结合地面反作用力数据 F_g、肢体段的重量和转动惯量，计算踝关节所受到的净作用力和力矩

如图 4.27 所示，假设某人体重为 25.2kg，需要计算其踝关节所受到的作用力和力矩。其步态周期中某个时刻的测量数据如下表所示。

	符号	单位	x_{lab}[⊖]	y_{lab}[⊖]	z_{lab}[⊖]
踝关节中心位置	A	m	0.357	0.823	0.056
脚尖标志点位置	T	m	0.421	0.819	0.051
压力中心位置	CP	m	0.422	0.816	0.000
地面反作用力矢量	$\boldsymbol{F}_g$	N	3.94	−15.21	242.36
地面反作用力力矩矢量	$\boldsymbol{T}_g$	N·m	0.000	0.000	0.995
脚的解剖坐标系	e_{fax}		0.977	−0.0624	−0.202
	e_{fay}		0.0815	0.993	0.0877
	e_{faz}		0.195	−0.102	0.975
脚的线加速度矢量	$\boldsymbol{a}_{foot}$	m/s^2	2.09	−0.357	−0.266
脚的角速度矢量	$\boldsymbol{\omega}_{foot}$	rad/s	0.0420	2.22	−0.585
脚的角加速度矢量	$\boldsymbol{\alpha}_{foot}$	rad/s^2	−0.937	8.85	−5.16
踝关节的角速度矢量	$\boldsymbol{\omega}_{ankle}$	rad/s	−0.000759	1.47	0.0106

下面利用本章表 4.1 所列的人体数据计算脚（foot）的质量和转动惯量及其重心的位置。脚的质量 m_{foot} 可以按照体重的 1.45%来估算，即为 0.365kg。重心的位置约在 50%脚长处。脚的长度 l_{foot} 可以根据踝关节中心到脚尖标志点之间的距离来估算，即

$$\begin{aligned}\boldsymbol{T}-\boldsymbol{A} &= (0.421-0.357)\boldsymbol{i}+(0.819-0.823)\boldsymbol{j}+(0.051-0.056)\boldsymbol{k}\\ &= 0.064\boldsymbol{i}-0.004\boldsymbol{j}-0.005\boldsymbol{k}\end{aligned}$$

$$\begin{aligned}l_{foot} &= |\boldsymbol{T}-\boldsymbol{A}|\\ &= \sqrt{(0.064)^2+(-0.004)^2+(-0.005)^2}=0.064\text{m}\end{aligned}$$

于是，重心相对于踝关节中心的位置是

⊖ x_{lab}、y_{lab}、z_{lab} 表示实验室固定坐标。——译者注

$$A+\frac{l_{foot}}{2}\frac{(T-A)}{|T-A|}=(0.357i+0.832j+0.056k)+\left[\frac{0.064}{2}\right]\left[\frac{0.064i-0.004j-0.005k}{0.064}\right]$$

即为

$$CG=0.389i+0.821j+0.054k$$

这样就可以计算图 4.27 所示的位置矢量 r_1 和 r_2。现在已知脚的长度为 0.064m，脚的质量为 0.365kg，旋转半径约为肢体段长度的 0.690，应用式（4.40）就可以计算相对于踝关节中心（ankle）的转动惯量，即

$$I_{foot/ankle}=(0.365kg)[(0.690)(0.064m)]^2$$
$$=7.12\times10^{-4}kg\cdot m^2$$

再利用式（4.41）表示的平行轴定理，就可以计算相对于脚的质心的转动惯量为

$$I_{foot/cm}=I_{foot/ankle}-m_{foot}d^2$$

注意，在均匀重力场作用下，质心与重心相同。式中 d 为脚的质心与踝关节之间的距离。由表 4.1 可知，脚的质心与其近端之间的距离为脚长的 0.5，因此，$d=0.5(l_{foot})=0.032m$，于是

$$I_{foot/cm}=(7.12\times10^{-4}kg\cdot m^2)-(0.365kg)(0.032m)^2$$
$$=3.38\times10^{-4}kg\cdot m^2$$

该值就是绕脚横截面上主轴 y' 和 z'（见图 4.27）的质心转动惯量，即

$$I_{y'y'}=3.38\times10^{-4}kg\cdot m^2$$
$$I_{z'z'}=3.38\times10^{-4}kg\cdot m^2$$

脚可以近似看作长度与半径之比为 6 的圆柱体，并且横向与纵向的转动惯量之比约为 6.5。因此，关于纵轴 x'（见图 4.27）的转动惯量就是

$$I_{x'x'}=5.20\times10^{-5}kg\cdot m^2$$

计算了这些脚的人体数据之后，下面就可以进行动力学分析。应用牛顿第二定律 $\Sigma F=ma$，可以计算踝关节所受到的反作用力 F_A，即

$$F_g+F_A-m_{foot}gk=m_{foot}a_{foot}$$

$$\begin{aligned}F_A&=m_{foot}a_{foot}-F_g+m_{foot}gk\\&=(0.365kg)[2.09i-0.357j-0.266k]m/s-\\&\quad(3.94i-15.21j+242.4k)N+\\&\quad(0.365kg)(9.81m/s^2)k\\&=-3.18i+15.08j-238.9k\ N\end{aligned}$$

然后，利用式（4.44）～式（4.46）表示的欧拉运动方程计算踝关节所受到的力矩 M_A。欧拉方程是相对于脚的主轴 x'、y' 和 z' 定义的，但是，要注意，这里需要计算的数据（如 ω_{foot} 和 α_{foot}）是相对于实验室坐标系（x，y，z）的数值。因此，首先要把欧拉方程求解所需的矢量转换到脚的解剖坐标系中。根据前面列表中的数据，脚的解剖坐标系为

$$e_{fax}=0.977i-0.0624j-0.202k$$
$$e_{fay}=0.0815i+0.993j+0.0877k$$

$$e_{faz}=0.195\boldsymbol{i}-0.102\boldsymbol{j}+0.975\boldsymbol{k}$$

式中 e_{fax}，e_{fay}，e_{faz}——分别对应于 x'、y'和 z'，即 $\boldsymbol{i}'$、$\boldsymbol{j}'$和 $\boldsymbol{k}'$。

在前面 4.2.2 节已经讲过，e_{fax}表达式中的系数分别是 x'对于 x、x'对于 y 和 x'对于 z 的余弦角；同理，e_{fay}表达式中的系数分别是 y'对于 x、y'对于 y 和 y'对于 z 的余弦角；e_{faz}表达式中的系数则分别是 z'对于 x、z'对于 y 和 z'对于 z 的余弦角。因此，将这三个表达式进行转置后，可得

$$\boldsymbol{i}=0.977\boldsymbol{i}'+0.0815\boldsymbol{j}'+0.195\boldsymbol{k}'$$
$$\boldsymbol{j}=-0.0624\boldsymbol{i}'+0.993\boldsymbol{j}'-0.102\boldsymbol{k}'$$
$$\boldsymbol{k}=-0.202\boldsymbol{i}'+0.0877\boldsymbol{j}'+0.975\boldsymbol{k}'$$

这样，利用这些关系式就可以将实验室坐标下的矢量 $\boldsymbol{A}=A_x\boldsymbol{i}+A_y\boldsymbol{j}+A_z\boldsymbol{k}$ 转换成脚坐标系下的矢量 $\boldsymbol{A}=A_x\boldsymbol{i}'+A_y\boldsymbol{j}'+A_z\boldsymbol{k}'$。为了说明这个方法，设脚的角速度矢量为

$$\boldsymbol{\omega}_{foot}=0.042\boldsymbol{i}+2.22\boldsymbol{j}-0.585\boldsymbol{k}\ \text{rad/s}$$

将脚坐标系表示的实验室坐标系代入，可得

$$\begin{aligned}\boldsymbol{\omega}_{foot}&=0.042(0.977\boldsymbol{i}'+0.0815\boldsymbol{j}'+0.195\boldsymbol{k}')+\\&2.22(-0.0624\boldsymbol{i}'+0.993\boldsymbol{j}'-0.102\boldsymbol{k}')-\\&0.585(-0.202\boldsymbol{i}'+0.0877\boldsymbol{j}'+0.975\boldsymbol{k}')\\&=0.0210\boldsymbol{i}'+2.16\boldsymbol{j}'-0.789\boldsymbol{k}'\text{rad/s}\end{aligned}$$

同样，可以将其他几个矢量变换到脚坐标系下，即

$$\begin{aligned}\boldsymbol{r}_1&=-0.032\boldsymbol{i}+0.002\boldsymbol{j}+0.002\boldsymbol{k}\\&=-0.032\boldsymbol{i}'-0.004\boldsymbol{k}'\text{m}\\\boldsymbol{r}_2&=0.033\boldsymbol{i}-0.005\boldsymbol{j}-0.054\boldsymbol{k}\\&=0.0435\boldsymbol{i}'-0.007\boldsymbol{j}'-0.0457\boldsymbol{k}'\text{m}\\\boldsymbol{F}_g&=3.94\boldsymbol{i}-15.21\boldsymbol{j}+242.36\boldsymbol{k}\\&=-44.16\boldsymbol{i}'+6.47\boldsymbol{j}'+238.62\boldsymbol{k}'\text{N}\\\boldsymbol{T}_g&=0.995\boldsymbol{k}\\&=-0.201\boldsymbol{i}'+0.0873\boldsymbol{j}'+0.970\boldsymbol{k}'\text{Nm}\\\boldsymbol{F}_A&=-3.18\boldsymbol{i}+15.1\boldsymbol{j}-239\boldsymbol{k}\\&=44.2\boldsymbol{i}'-6.23\boldsymbol{j}'-235\boldsymbol{k}'\text{N}\\\boldsymbol{\omega}_{foot}&=0.0420\boldsymbol{i}+2.22\boldsymbol{j}-0.585\boldsymbol{k}\\&=0.021\boldsymbol{i}'+2.16\boldsymbol{j}'-0.789\boldsymbol{k}'\text{rad/s}\\\boldsymbol{\alpha}_{foot}&=-0.937\boldsymbol{i}+8.85\boldsymbol{j}-5.16\boldsymbol{k}\\&=-0.425\boldsymbol{i}'+8.26\boldsymbol{j}'-6.11\boldsymbol{k}'\text{rad/s}^2\end{aligned}$$

根据式（4.44）～（4.46）表示的欧拉运动方程，有

$$\boldsymbol{M}_{Ax'}+(r_1\times\boldsymbol{F}_A)_{x'}+(r_2\times F_g)_{x'}+\boldsymbol{T}_{gx'}=\boldsymbol{I}_{x'x'}\boldsymbol{\alpha}_{x'}+(\boldsymbol{I}_{z'z'}-\boldsymbol{I}_{y'y'})\boldsymbol{\omega}_{y'}\boldsymbol{\omega}_{z'}$$
$$\boldsymbol{M}_{Ay'}+(r_1\times\boldsymbol{F}_A)_{y'}+(r_2\times F_g)_{y'}+\boldsymbol{T}_{gy'}=\boldsymbol{I}_{y'y'}\boldsymbol{\alpha}_{y'}+(\boldsymbol{I}_{x'x'}-\boldsymbol{I}_{z'z'})\boldsymbol{\omega}_{z'}\boldsymbol{\omega}_{x'}$$
$$\boldsymbol{M}_{Az'}+(r_1\times\boldsymbol{F}_A)_{z'}+(r_2\times F_g)_{z'}+\boldsymbol{T}_{gz'}=\boldsymbol{I}_{z'z'}\boldsymbol{\alpha}_{z'}+(\boldsymbol{I}_{y'y'}-\boldsymbol{I}_{x'x'})\boldsymbol{\omega}_{x'}\boldsymbol{\omega}_{y'}$$

式中 $(\boldsymbol{r}_1\times\boldsymbol{F}_A)_{x'}$——$\boldsymbol{r}_1\times\boldsymbol{F}_A$的 x'分量；

$(r_2 \times F_g)_{x'}$——$r_2 \times F_g$的 x'分量，以此类推。

将所需的数值代入，并简化，可得

$$\boldsymbol{M}_{A'} = 1.50\boldsymbol{i}' + 15.9\boldsymbol{j}' - 1.16\boldsymbol{k}'\text{Nm}$$

再将其转换到实验室固定坐标系下，即

$$\boldsymbol{M}_{A} = 2.54\boldsymbol{i} + 15.9\boldsymbol{j} - 0.037\boldsymbol{k}\text{Nm}$$

将踝关节力矩与踝关节角速度相乘，可得踝关节功率

$$\begin{aligned}\boldsymbol{M}_{A} \cdot \boldsymbol{\omega}_{\text{ankle}} &= (2.54\boldsymbol{i} + 15.9\boldsymbol{j} - 0.037\boldsymbol{k}\ \text{Nm}) \cdot (-0.000759\boldsymbol{i} + 1.47\boldsymbol{j} + 0.0106\boldsymbol{k}\ \text{rad/s}) \\ &= 23.3\text{W}\end{aligned}$$

即

$$\begin{aligned}\boldsymbol{M}_{A'} \cdot \boldsymbol{\omega}_{\text{ankle}'} &= (1.50\boldsymbol{i}' + 15.9\boldsymbol{j}' - 1.16\boldsymbol{k}'\ \text{Nm}) \cdot (-0.0946\boldsymbol{i}' + 1.46\boldsymbol{j}' - 0.140\boldsymbol{k}'\ \text{rad/s}) \\ &= 23.3\text{W}\end{aligned}$$

该数值就是踝关节所贡献的步行推进力的定量测定。

4.6.4 临床步态数据的含义

临床步态分析中，用于确定治疗方案的信息和数据不仅包括上述定量数据，即躯干、骨盆、髋关节、膝关节、踝关节和脚等肢体段的角位移等三维运动学数据，以及髋关节、膝关节和踝关节的力矩和功率等三维动力学数据，还包括以下信息：

1）临床体检数据；

2）病人行走的双平面视频记录数据；

3）步幅长度和行走速度等空间和时间上的步态数据；

4）某些下肢肌肉的肌电图记录（EMG）。

一般，步态数据的分析工作包括畸变的识别、畸变病因的确定、治疗方案的推荐等。结合多方面的数据，通过分析各组数据之间的相关性，化解数据中明显的冲突，就可以形成病人步行能力的正确诊断。病人自身的补偿机制会在某种程度上帮助病人克服步行障碍，医疗组人员根据经验可以将导致病人步行困难的步态异常与补偿机制区分开来。

下面举例说明这些分析过程。图 4.28～图 4.30 所示是一位患有双侧痉挛性脑瘫的 9 岁女孩的步态测量数据。脑性瘫痪（Cerebral palsy）是一种非进行性脑损伤引起的神经肌肉疾病，常由出生时或出生不久发生的脑损伤引起。最常见的是大脑皮层运动神经受损，使神经对于肌肉的控制减弱，引起肢体各组肌肉运动的不协调或不平衡，如活动过度、活动失调和肌肉痉挛等。治疗方法有理疗、配戴支具等矫形器、服用肉毒毒素（botulinus toxin）和巴氯芬（Baclofen）等缓解痉挛的药物以及施行整形手术和神经外科手术等。

图 4.28 所示的病人矢状面左侧运动学数据显示了其髋关节和膝关节的重要信息。在支撑期（stance phase），即步态周期的 0～60%，在脚接触地面的整个过程中，她的膝关节被有力地“锁”在一个过度弯曲的位置；在摆动期（swing phase），即步态周期的 60%～100%，膝关节的活动仍然受到限制，摆动时膝关节弯曲的最大幅度

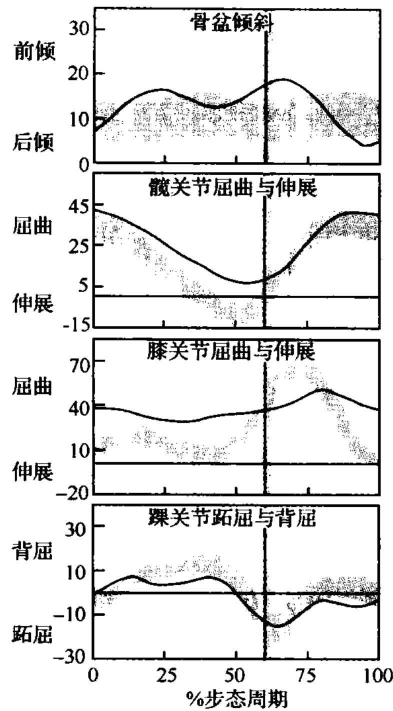


图 4.28 某 9 岁双侧痉挛性脑瘫病人的矢状面左侧运动学数据（实线）

注：灰色条带表示正常儿童步态数据的正、负一个标准差范围。图中垂直实线表示步态周期的 0～60%是支撑期，60%～100%是摆动期。

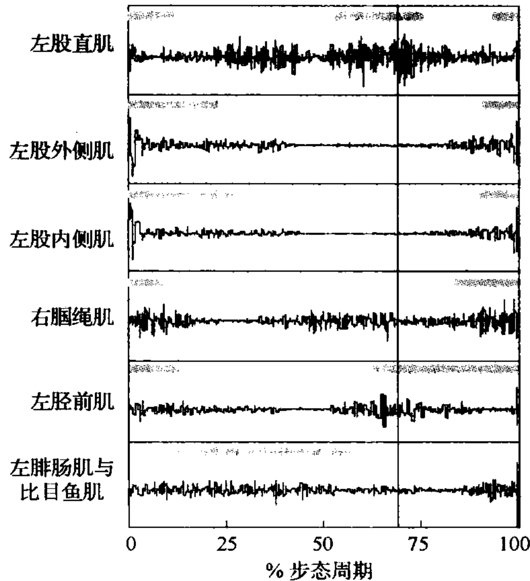


图 4.29 图 4.28 所示脑瘫病人的肌电图

注：图中所示分别是左下肢 6 块肌肉的肌电图信号。横坐标是步态周期。灰色条带表示正常肌肉的平均激活时间。

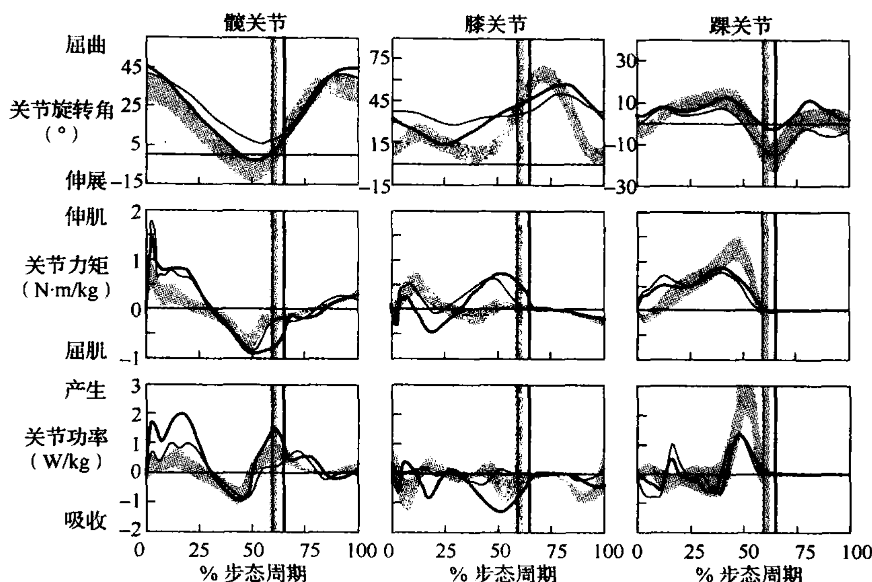


图 4.30 图 4.28 所示脑瘫病人的矢状面关节动力学数据

注：图中分别是髋关节、膝关节和踝关节的旋转角、力矩和功率随步态周期变化的曲线。粗、细实线分别表示右侧和左侧数据。灰色条带表示正常人群步态数据的正、负一个标准差范围。

减小，弯曲时间也推迟。其髋关节的活动也小于正常幅度，在支撑期的末期大腿没能完全伸展。而骨盆的活动幅度却明显大于正常人，支撑期早期的骨盆前倾与大腿后伸同时发生，摆动期的骨盆后倾与大腿前屈同时发生。

这些数据偏离正常状态的情况显示了这类病人常见的神经肌肉障碍。临床体检中观察到不协调的腿腱紧绷以及图 4.29 肌电图所示的支撑期的肌肉活动不协调，使得膝关节不能正常伸展。这种膝关节弯曲的姿势也阻碍了大腿的正常伸展（临床体检中也观察到了大腿伸肌衰弱的现象）。步行时的支撑期需要大腿后伸，这样在骨盆和上身前移时大腿就可以旋转。为了弥补大腿后伸的不足，她在支撑期的早期就旋转骨盆，以增加大腿的移动幅度。骨盆倾斜角曲线上的两个波峰表明在一定程度上双侧都存在这个问题。

在摆动期，膝关节弯曲受限，再加上踝关节的跖屈姿势，使得足部离地困难。图 4.29 肌电图显示的股直肌在摆动期中期的过度活动说明该肌肉处于痉挛状态，股直肌是膝关节伸肌之一，它的痉挛阻碍了膝关节的弯曲。而且，摆动期后期踝关节的跖屈肌（主要是腓肠肌）的活动增强，超过了胫前肌，导致踝关节处于跖屈位，形成足下垂畸形。

图 4.30 所示的该病人矢状面关节动力学数据显示了左右两侧活动的不对称。值得注意的是，她在利用右膝关节和右大腿补偿左侧的某些功能障碍。特别是从支撑期开始逐渐增加的右膝关节弯曲（图中第一行中间）一直延续到摆动期，这种弯曲有助于对侧肢体在摆动期的前移。这种补偿的一个潜在不利后果是支撑期后期膝关

节伸肌的力矩加大，从而增加了髌骨一股骨的负荷，其长期的作用结果难以预料。髋关节处产生的不对称的功率作用也显而易见，右下肢（特别是连接在右侧髋部上的肌肉）在支撑期早期产生的功率很大，它提供的力拖动身体前移并抬高重心。而且，在足尖离地时髋部大而快的功率发生加速了肢体进入摆动期，并且，不顾膝关节伸肌力矩的增加，促使膝关节弯曲。重要的是要注意双侧跖屈肌的痉挛反应，踝关节的过早发力以及临床牵张反射检查中的痉挛现象都可证实这一点。女孩向前行走时对大腿部位肌肉的利用远远超过踝关节的跖屈肌，这一点右侧比左侧更明显。

从这个简略的病例分析可见，步态分析可以识别病人步行的异常，揭示其中的生物力学病因。上述分析中讨论的是神经肌肉疾病对于病人矢状面上步态数据的某些影响。临床步态分析还可以通过有关静态骨骼旋转畸形资料来揭示步态异常。另外，考察支具、手术等的治疗效果对于临床研究也很有用。不过，应该注意，虽然工程师和应用物理学家早在 100 年之前就已经开始加入这个领域，但是，目前这个领域仍然存在重大机遇，临床步态分析的生物力学方法和工具仍需改进，许多东西还需要学习。

4.7 心血管动力学

心血管系统的动力学是力学在人体器官系统中的一个应用，也称为血液动力学（hemodynamics），它研究血液的流动。从功能上看，心血管系统由复杂的泵（即心脏）、复杂的液体（即血液）和复杂的管道网络（即血管）构成。心脏产生血压，使血液能够流过血管。心血管动力学主要研究心血管系统的血压、容积以及血流的测量和计算。心血管系统非常复杂，很精细，要用各种力学模型，特别是数学方程表示的模型，才能分析和综合实验数据，剖析和认识其中的生理机制，最终开发出新的心血管功能临床检测方法，用于指导临床治疗。

本书第 3 章已经介绍过，心脏是一个包含 4 个腔室的泵，它与体循环和肺循环两套血管系统相连。心泵的运行由电活动触发，受到神经系统和激素系统的调控。心脏的单向阀门（即瓣膜）控制血流方向。人体的总血容量约为 5.2L。左心室是心脏最强的腔室，静息状态下每分钟能够泵血 5L 左右，几乎接近人体血液的总量。心脏每搏动一次，左心室泵血约 70ml，每分钟平均心搏约 72 次。体育锻炼时，左心室的输出量可以增加到静息时的 6 倍，心率可以超过静息时的 2 倍。整个血液循环系统的血管总长约为 100000km，大约可以绕地球 2.5 周。静息时左心室产生的机械功率约为 1.7W，剧烈运动时可以增加 3 倍。还有一个有趣的数字是哺乳动物一生的心搏总数大约为十亿次（Vogel，1992）。大动物的心率较慢，寿命较长；而小动物心率快，寿命也较短。

4.7.1 血液流变学

血液由血浆（plasma）以及其中的细胞组成。血液中的细胞包括红细胞（eryth-

rocyte)、白细胞（leukocyte）和血小板（platelet）。如图4.31所示，流体与固体在力学上的区别是：固体能够通过内部产生的反作用力（图中虚线箭头表示）抗衡外加的两个平行且反向的剪切作用力（图中实线箭头表示）；而流体却不能抗衡这种剪切力的作用，只能流动。

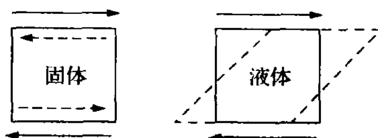


图4.31 左图：固体材料通过内部产生的反作用剪应力（虚线箭头）来抗衡外加的剪切力（实线箭头）；右图：流体不能抗衡外加剪切力，只能流动（虚线所示）

外加的剪切力会产生剪应力（即单位面积上的作用力），流体在剪应力作用下的流动可以用剪应变率（shear strain rate）定量描述。一般，流体受到的推力越大，也就是剪应力越大，流体流动就越快，剪应变率也越高。剪应力 τ 与剪应变率 $\dot{\gamma}$ 之间的关系用流体的粘度μ（viscosity）来表示，在生物医学应用中常将粘度记为 η。如图4.32所示，水和空气等很多流体的粘度都是常数，具有这种线性特性的流体被称为牛顿流体。具有非线性剪应力与剪应变率关系的流体则称为非牛顿流体。例如，粘度会随着剪应变率的增加而增加的流体称为胀流型流体（dilatant），它属于非牛顿流体。Dow Corning 3179 化合物就是一种胀流型流体，它是硅聚合物，俗称“橡皮泥”。如果慢慢地拉伸橡皮泥，它就会伸长，产生塑性变形；如果快速拉它，那么，橡皮泥就会像固体一样断裂。另外，剪应变率较大、粘度较小的流体被称为假塑性流体（pseudo-plastic）。例如，防滴乳胶涂料用刷子或滚筒涂刷时（即剪应力作用下）会流动，但涂完之后就不流动了。

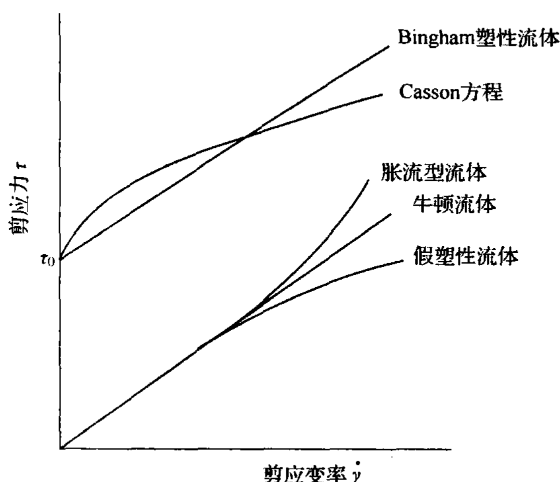


图4.32 流体的剪应力与剪应变率关系

注：牛顿流体的粘度为常数，其剪应力与剪应变率之间呈线性关系。而非牛顿流体的剪应力与剪应变率之间呈非线性关系。血液的特性通常用 Casson 方程描述，不过在许多情况下也用牛顿流体来描述。

生物流体一般都是非牛顿流体。但血浆属于牛顿流体，其物理特性与水很相似。全血则像 Bingham 塑性流体，它在某个大于屈服应力（yield stress）的非零剪应力作用下才会开始流动。血液的特性常用如下幂律函数表示：

$$\tau = k\dot{\gamma}^n \tag{4.57}$$

式中 k 和 n 为常数，它们可以通过 $\ln\tau$ 与 $\ln\dot{\gamma}$ 之间的线性拟合来确定，即

$$\ln\tau = \ln k + n\ln\dot{\gamma}$$

另一个常用的血液粘度方程是 Casson 方程：

$$\tau^{\frac{1}{2}} = \tau_0^{\frac{1}{2}} + k\dot{\gamma}^{\frac{1}{2}} \tag{4.58}$$

由 Casson 曲线可以测得血液的屈服应力 τ_0。流变学主要研究具有复杂粘性的流体的变形和流动。有关生物流体流变学的教科书有 Mazumdar（1992）的著作《生物流体力学》（Biofluid Mechanics）和 Fournier（1999）的著作《生物医学工程中的基本传输现象》(Basic Transport Phenomena in Biomedical Engineering)。

例 4.12 以下是某个血液样本的流变学测量数据：

剪应变率/s^{-1}	剪应力/（$dyne/cm^2$）
1.5	12.5
2.0	16.0
3.2	25.2
6.5	40.0
11.5	62.0
16.0	80.5
25.0	120
50.0	140
100	475

请编写 MATLAB 程序，用一个幂律函数拟合这些数据。

解：MATLAB 程序如下：

```
% Power Law Fit of Blood Data
%
% Store shear strain rate and stress data in arrays
alpha = [1.5,2,3.2,6.5,11.5,16,25,50,100] ;
T = [12.5,16,25.2,40,62,80.5,120,240,475] ;
% Take natural logs of both
x = log(alpha) ;
y = log(T) ;
% Use MATLAB's polyfit function to do linear curve fit
coeff = polyfit (x,y,1)
% Write curve fit coefficients as a new x-y function for plotting
x1=[0;0.01;5]
y1=polyval (coeff,x1)
% Plot the original data as 'o' points
plot(x,y, 'o')
hold on
% Overlay a plot of the curve-fit line
plot(x1,y1)
grid on
```

```
title ('Power Law Function')
xlabel('ln Strain Rate [ln (1/s)]')
ylabel('ln Shear Stress [ln dyne/cm2]')
%
```

该程序的输出结果如图 4. 33 所示。

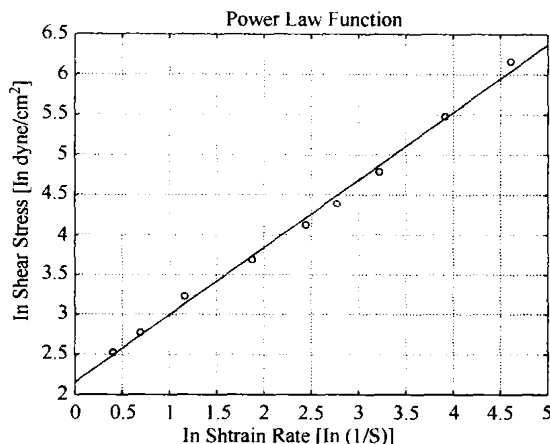


图 4. 33 例 4. 12 的 MATLAB 程序的输出结果（利用幂律函数拟合血液流变学数据） ■

当剪应变率很小时，血液所呈现的粘度要比预期的大，这是红细胞聚集的结果，这种聚集称为红细胞缗钱叠加（rouleaux），这时的剪应变率比一般大血管或者医学仪器中的剪应变率都要小。当剪应变率较高时，直径小于 1mm 的小血管中血液所呈现的粘度要比大血管中的粘度小，这种现象被称为 Fahraeus-Lindquist 效应，它是由血浆与红细胞之间的动力学作用造成的。除了这两种特殊情况以外，血液可以看作牛顿流体，这已被科学界广泛认可。下面我们将看到牛顿流体的假设可以大大简化人体循环系统的力学分析。

4. 7. 2 动脉血管

血管力学特性的研究具有悠长而复杂的历史，其中所涉及的多数高等数学和应用力学的知识都超出了本书的范围。本节将只概述其主要的发展过程，并介绍本章后面将用到的一种简化的动脉系统模型。有关更深入的内容，请读者参考下列教科书：Noordergraaf（1978）的著作《循环系统动力学》（Circulatory System Dynamics）、Milnor（1989）的著作《血液动力学》（Hemodynamics）和 Mazumdar（1992）的著作《生物流体力学》（Biofluid Mechanics），有关基础流体力学的著作有 White（2003）的《流体力学》(Fluid Mechanics)。

为了研究心脏泵血功能的力学特性，需要计算心肌作用力和心肌活动引起的血压和血流；因此，心血管系统的通用运动方程一般都利用线动量守恒定律建立。将 Reynold 传输定理（即流体力学的守恒方程）应用于线性动量就可以产生如下适用于

任何流体的通用运动方程：

$$\rho \mathrm{g}-\nabla p+\nabla \cdot \tau_{ij}=\rho \frac{\mathrm{d}V}{\mathrm{d}t} \tag{4.59}$$

式中 ρ——流体密度，即质量/容积；

g——重力加速度；

p——压力；

τ_{ij}——粘性力；

V——速度；

∇——微分算子，即$\nabla=i\frac{\partial}{\partial x}+j\frac{\partial}{\partial y}+k\frac{\partial}{\partial z}$。

速度矢量V是空间位置和时间的函数，可以写为

$$V(x,y,z,t)=u(x,y,z,t)i+v(x,y,z,t)j+w(x,y,z,t)k$$

式中 u、v和w——分别为x、y和z方向的局部速度。

方程（4.59）由重力、压力、粘性力和时间变化4部分组成。注意，它是一个矢量方程，可以展开成如下x、y和z三个分量的联立方程：

$$\rho g_x-\frac{\partial p}{\partial x}+\frac{\partial \tau_{xx}}{\partial x}+\frac{\partial \tau_{yx}}{\partial y}+\frac{\partial \tau_{zx}}{\partial z}=\rho\left(\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}+w\frac{\partial u}{\partial z}\right) \tag{4.60}$$

$$\rho g_y-\frac{\partial p}{\partial y}+\frac{\partial \tau_{xy}}{\partial x}+\frac{\partial \tau_{yy}}{\partial y}+\frac{\partial \tau_{zy}}{\partial z}=\rho\left(\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}+w\frac{\partial v}{\partial z}\right) \tag{4.61}$$

$$\rho g_z-\frac{\partial p}{\partial z}+\frac{\partial \tau_{xz}}{\partial x}+\frac{\partial \tau_{yz}}{\partial y}+\frac{\partial \tau_{zz}}{\partial z}=\rho\left(\frac{\partial w}{\partial t}+u\frac{\partial w}{\partial x}+v\frac{\partial w}{\partial y}+w\frac{\partial w}{\partial z}\right) \tag{4.62}$$

该非线性偏微分方程组是个通用公式，但它不能直接求解，简化后才能求解。例如，如果忽略流体的粘性力，式（4.59）就可以简化为无粘性流体的欧拉方程。如果再将欧拉方程沿流线积分，就产生了著名的柏努利方程（Bernoulli equation），该方程描述了压力与流量之间的关系。在实际应用中，柏努利方程的无粘性条件（也就是无摩擦力）经常会被遗忘。

如果流体处于稳态流动状态，那么，式（4.59）右边的项就等于0；另外，对于不可压缩流体，密度ρ为常数，这样，就可以大大简化包含ρ的重力项和时变项的积分。同理，对于牛顿流体，粘度μ为常数，也可以简化方程。总之，虽然可以写出完整的通用流体运动方程，但是，求解这些方程很困难，必须进行某些合理的简化。

大血管中的血流可以假设为牛顿流体，并且是不可压缩流体。有了这两个假设，式（4.59）就可以简化为如下Navier—Stokes方程：

$$\rho g_x-\frac{\partial p}{\partial x}+u\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}\right)=\rho\frac{\mathrm{d}u}{\mathrm{d}t} \tag{4.63}$$

$$\rho g_y-\frac{\partial p}{\partial y}+u\left(\frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}+\frac{\partial^2 v}{\partial z^2}\right)=\rho\frac{\mathrm{d}v}{\mathrm{d}t} \tag{4.64}$$

$$\rho g_z-\frac{\partial p}{\partial z}+u\left(\frac{\partial^2 w}{\partial x^2}+\frac{\partial^2 w}{\partial y^2}+\frac{\partial^2 w}{\partial z^2}\right)=\rho\frac{\mathrm{d}w}{\mathrm{d}t} \tag{4.65}$$

对于血管，用柱坐标比用直角坐标方便，因此，可以将上式的坐标x、y和z转化为

柱坐标系的半径 r、角度 θ 和轴向距离 x。如果流体为无旋流体，即 $\theta=0$，那么，只要用两个 Navier—Stokes 方程就足够了，即

$$-\frac{\mathrm{d}p}{\mathrm{d}x}=\rho\left(\frac{\mathrm{d}w}{\mathrm{d}t}+u\frac{\mathrm{d}w}{\mathrm{d}r}+w\frac{\mathrm{d}w}{\mathrm{d}x}\right)-u\left(\frac{\mathrm{d}^2w}{\mathrm{d}r^2}+\frac{1}{r}\frac{\mathrm{d}w}{\mathrm{d}r}+\frac{\mathrm{d}^2w}{\mathrm{d}x^2}\right) \tag{4.66}$$

$$-\frac{\mathrm{d}p}{\mathrm{d}r}=\rho\left(\frac{\mathrm{d}u}{\mathrm{d}t}+u\frac{\mathrm{d}u}{\mathrm{d}r}+w\frac{\mathrm{d}u}{\mathrm{d}x}\right)-u\left(\frac{\mathrm{d}^2u}{\mathrm{d}r^2}+\frac{1}{r}\frac{\mathrm{d}u}{\mathrm{d}r}+\frac{\mathrm{d}^2u}{\mathrm{d}x^2}-\frac{u}{r^2}\right) \tag{4.67}$$

式中　w——纵向速度 $\mathrm{d}x/\mathrm{d}t$；

　　　u——径向速度 $\mathrm{d}r/\mathrm{d}t$。

一般动脉血管模型还使用源自质量守恒定律的连续方程，即

$$\frac{\mathrm{d}u}{\mathrm{d}r}+\frac{u}{r}+\frac{\mathrm{d}w}{\mathrm{d}x}=0 \tag{4.68}$$

该方程表明：系统内储存质量的净变化率等于流入质量的净变化率减去流出质量的净变化率。

Noordergraaf 和同事们（1969）将式（4.66）的 Navier-Stokes 方程改写为

$$-\frac{\mathrm{d}P}{\mathrm{d}x}=RQ+L\frac{\mathrm{d}Q}{\mathrm{d}t} \tag{4.69}$$

式中　P——压力；

　　　Q——容积流速；

　　　R——等效阻抗；

　　　L——流体惯量。

Navier-Stokes 方程描述了血管内血液的流体力学。由于动脉血管壁具有弹性，因此，血管壁也需要用运动方程来描述。描述血管壁的方程从简单发展到复杂，已经有线性弹性方程、线性粘弹性方程和复杂粘弹性方程等（参见 Noordergraaf，1978）。通用性最强的各向异性的线性动脉血管壁的力学方程用了 21 个参数（参见冯元桢，1977），其中多数参数从来没有测量过。Noordergraaf 等人将动脉血管分成许多小段，并将各小段血管的流体力学方程（即式（4.69））与连续方程结合起来。动脉血管壁的弹性使得血管会随着心脏的搏动产生扩张，从而导致血管内储存的血流量随时间发生变化。小段动脉血管的连续方程为

$$-\frac{\mathrm{d}Q}{\mathrm{d}x}=GP+C\frac{\mathrm{d}P}{\mathrm{d}t} \tag{4.70}$$

式中　G——血管壁的渗漏。

Noordergraaf 等人用式（4.69）和式（4.70）这对方程描述了 125 个小段组成的动脉系统，这是第一个充分描述动脉压和血流反射波的模型。在心血管脉动系统中，动脉分支导致了血压和血流的反射波。物理上的 RLC 电路及其连接形成的大输电网络线路中的电压和电流就好比是液流系统的压力和流量。如果不考虑脉动波反射等分布式动脉特性，那么，从心脏看出去的动脉系统的负荷就可以大为简化，就像电网可以简化成等效电路一样。

图 4.34 所示的三元件模型是应用最广泛的动脉负荷模型。由于该模型早在数字

计算机发明之前就已经建立，因此，它是一个电路模型。其中，Z_0表示主动脉的阻抗特性，实际上是主动脉血流阻力；C_s是动脉的横向顺应性，即弹性系数的倒数，用于描述动脉系统的径向扩张能力，它相当于电容；R_s是外周阻抗，用于描述主动脉下游循环系统动脉的血流阻力。这个简单电路模型描述了从左心室看出去的动脉系统的负荷。根据这个模型可以写出如下常微分方程，其左边表示与血压 $p(t)$ 相关的项，右边则表示与血流 $Q(t)$ 相关的项，它描述了 $p(t)$ 与 $Q(t)$ 之间的关系：

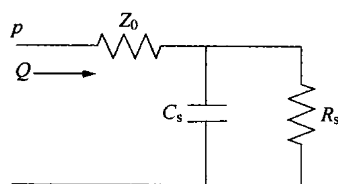


图 4.34　动脉系统负荷的等效电路模型（各个电路元件的含义见正文所述）

$$C_s\frac{\mathrm{d}p}{\mathrm{d}t}+\frac{1}{R_s}p(t)=Q(t)\left(1+\frac{Z_0}{R_s}\right)+Z_0C_s\frac{\mathrm{d}Q}{\mathrm{d}t} \tag{4.71}$$

例 4.13　请应用基本电路理论，从图 4.34 推导出微分方程（4.71）。

解：如图 4.35 所示，根据基尔霍夫电流定律，可以写出电路中节点 1 的电流方程

$$Q=Q_1+Q_2$$

而 Q_1和 Q_2分别为

$$Q_1=C_s\frac{\mathrm{d}p_1}{\mathrm{d}t}$$

$$Q_2=\frac{p_1}{R_s}$$

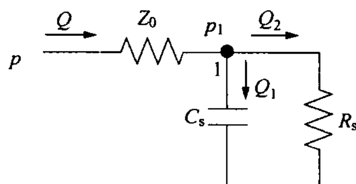


图 4.35　三元件动脉系统负荷的电路节点分析

因此

$$Q=C_s\frac{\mathrm{d}p_1}{\mathrm{d}t}+\frac{p_1}{R_s}$$

根据欧姆定律，有

$$p-p_1=QZ_0$$

由上式求得 p_1，并将其代入 Q 的表达式，可得

$$\begin{aligned}Q&=C_s\frac{\mathrm{d}}{\mathrm{d}t}[p-QZ_0]+\frac{1}{R_s}[p-QZ_0]\\&=C_s\frac{\mathrm{d}p}{\mathrm{d}t}-Z_0C_s\frac{\mathrm{d}Q}{\mathrm{d}t}+\frac{1}{R_s}p-\frac{Z_0}{R_s}Q\end{aligned}$$

将 Q 的项集中到等式右边，并将 p 的项放到等式左边，就得到式（4.71）。■

4.7.3　心脏力学

心脏的力学特性，特别是左心室，通常用心室的弹性系数（elastance）来描述。心脏就像一个有弹性的口袋，每次心跳都要收缩和舒张。弹性系数是衡量物体刚性的一个指标，定义为压力对于容积的微分，即：

$$E_v=\frac{\mathrm{d}p_v}{\mathrm{d}V_v} \tag{4.72}$$

式中　p_v——心室血压；

V_v——心室容积；

E_v——心室的时变弹性系数。

可见，E_v是血压相对于容积的微分变化。这种关系在数学上很明确，但是E_v的实际测量却很困难。

20世纪70年代，Kennish等人用以下心室血压与容积的有限变化量之比来估算式（4.72）的微分：

$$E_v=\frac{\Delta p_v}{\Delta V_v} \tag{4.73}$$

不过，这种算法会产生实际不可能出现的结果。例如，在动脉瓣打开之前的心脏等容收缩期，左心室血压不断上升，而此时容积却没有任何变化，由于式（4.73）的分母为0，得到的弹性系数E_v就变成无穷大了。为了解决这个问题，Suga和Sagawa（1974）不用微分也不用差分，而是直接用血压与容积的比来估算弹性系数，即

$$E_v(t)=\frac{p_v(t)}{V_v(t)-V_d} \tag{4.74}$$

式中　V_d——容积常数，它指固定不变的死腔容积。此式中的其他项都可以随时间变化。

应用这种算法，从实验数据计算得到的心室弹性曲线如图4.36所示，图中的误差线表明曲线的方差很大。显然，图中所示E_v曲线的形状明显不对称，这与等容心搏实验的结果相矛盾。等容心搏是一种简化实验，实验时将主动脉结扎，不让左心室射血，从而形成等容心搏状态。由式（4.74）可知，在等容情况下（即V_v为常数），心室血压p_v的波形必定与弹性系数E_v相同．但是，该实验获得的等容血压曲线都是对称的，不像图4.36所示的E_v的非对称曲线。其原因在于具有射血心搏的$E_v(t)$不仅与心脏（即心室）的特性相关，还与循环系统（即血管）有关。因此，图4.36的时变弹性曲线实际上反映的是心脏（源）与血管系统（负荷）共同作用的结果，除了心室特性以外，它还受到血管特性的影响。而式（4.74）表示的算法并不

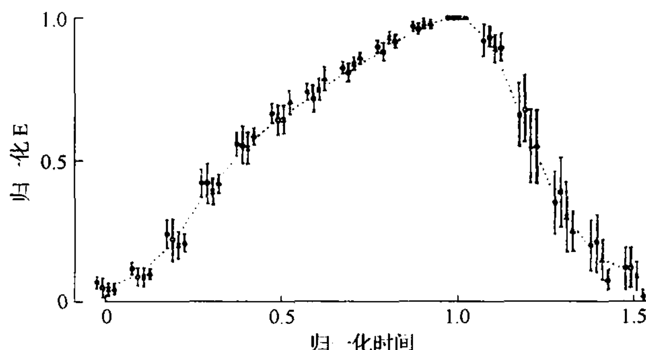


图4.36　根据式（4.74）的定义测量并计算的心室时变弹性系数曲线，图中所示弹性曲线形状不对称（摘自Suga和Sagawa，1974）

能单独将心室特性与血管特性分离开来。因此，需要设计新的描述心脏力学特性的方法。

为了解决上述等容特性与射血特性的不一致，以及心脏与血管的耦合特性等问题，有人开发了一种新的左心室力学特性描述方法（Mulier，1994；Palladino 等人，1997）。这种模型既简单又通用，并且与仅仅模拟生理行为的那些仿真不同，这个模型具有直接的生理学意义。如图 4.37 所示，该模型根据狗的离体心脏实验数据建立。实验开始时先将左心室充灌不同容积的初始血液，形成各种不同的负荷状态，然后刺激心脏，使其搏动。同时，测量并记录心室血压，某些实验中还测量心室流量。

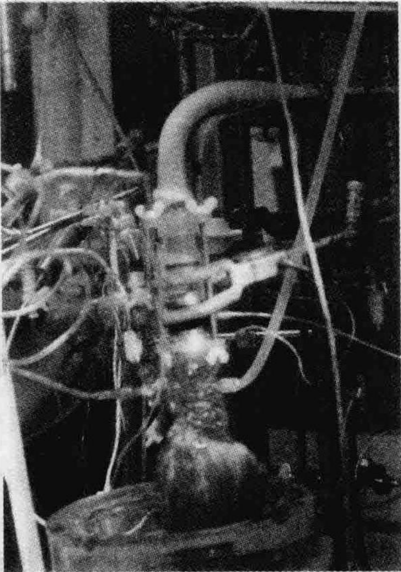


图 4.37 用于建立心脏生物力学新模型的狗离体心脏（照片由比利时 Leuven 的 Jan Mulier 博士馈赠）

实验中测定等容心室血压曲线时，每次先将离体左心室充灌一定的初始血容量（即舒张末期容量），然后夹住主动脉，使血液不能流出。再刺激心脏，使其搏动，并检测和记录心室所产生的血压。充灌不同的初始血容量，重复上述实验过程。就像 Otto Frank（1895）所做的著名实验一样，此时心室的等容血压直接与血液充灌量相关。图 4.38 显示了正常狗左心室测得的一组等容血压曲线。

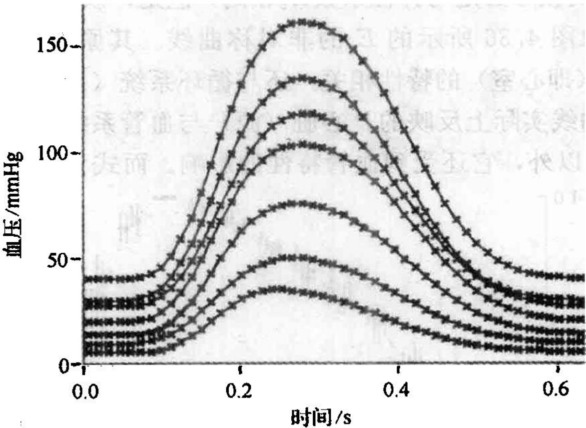


图 4.38 等容心室血压曲线

注：每一条曲线的记录过程是：左心室充灌一定的初始血容量之后，夹住主动脉，然后刺激心脏，并测量心室血压随时间的变化。图中每条曲线对应于不同的初始血容量。

这些等容血压曲线可以用如下方程描述，其中，心室血压 p_v 是时间 t 和心室容

积 V_v的函数，即

$$p_v=a(V_v-b)^2+(cV_v-d)\left[\frac{(1-\mathrm{e}^{-\left(\frac{t}{\tau_c}\right)})\mathrm{e}^{-\left(\frac{t-t_b}{\tau_r}\right)^{\alpha}}}{(1-\mathrm{e}^{-\left(\frac{t_p}{\tau_c}\right)^{\alpha}})\mathrm{e}^{-\left(\frac{t_p-t_b}{\tau_r}\right)^{\alpha}}}\right] \tag{4.75}$$

该方程可以简写为

$$p_v(t,V_v)=a(V_v-b)^2+(cV_v-d)f(t) \tag{4.76}$$

式中　$f(t)$ ——式（4.75）方括号中的激活函数；

a、b、c、d、t_p、t_b、τ_c、τ_r和 α——常数，由离体狗心室实验测得。

从生理学上看，式（4.76）将心室描述成一个压力发生器，其产生的压力随时间和容积而变化。其中，包含常数 a 和 b 的项描述了心室的被动弹性特性，包含常数 c 和 d 的项则描述了由心肌作用力所产生的心室的主动弹性特性。表 4.3 列出了狗心室实验中测得的该模型的典型参数，表中右边还列出了经过修正之后用于描述人体左心室模型的参数（Palladino 等人，1997）。

表 4.3　动物实验中测得的心室模型参数以及修正之后用于人体心脏分析的模型参数

参　数	狗（实验测得）	人（修正所得）
a/（mmHg/ml²）	0.003	0.0007
b/ml	1.0	20.0
c/（mmHg/ml）	3.0	2.5
d/mmHg	20.0	80.0
τ_c/s	0.164	0.264
t_p/s	0.271	0.371
τ_r/s	0.199	0.299
t_b/s	0.233	0.258
α	2.88	2.88

例 4.14　请编写 MATLAB 程序，求解式（4.75），并画出人体心脏一个搏动周期的心室血压曲线 p_v（t）。假设初始心室容积为 150ml，其他参数值见表 4.3。

解： 以下 MATLAB 程序可以完成所需计算，并作出图 4.39 所示的曲线。

```
% ventricle.m
%
% MATLAB m-file to compute isovolumic pressure using Mulier ventricle model
%
% Initial conditions:
%
delt = 0.001; % The interation time step delta t
a = 7e-4;
b = 20.;
c = 2.5;
d = 80.;
tc = 0.264;
tp = 0.371;
```

```
tr = 0.299;
tb = 0.258;
alpha = 2.88;
Vv0 = 150; % Initial (end-diastolic) ventricular volume
%
% Compute an intermediate term denom
% to simplify computations:
%
  denom = ((1. - exp ( - (tp/tc)^alpha) ) * exp ( - ( (tp - tb)/tr)^alpha) );
%
% Compute for initial time t = 0 (MATLAB does not allow 0 index)
%
  t (1) = 0.;
  Vv (1) = Vv0;
  edp = a * ( (Vv0 - b) )^2;
  pdp = c * Vv0 - d;
  pp = pdp/denom;
  t1 = 0.; % Time step for first exponential
  t2 = 0.; % Time step for second exponential
  e1 = exp ( - (t1/tc)^alpha);
  e2 = exp ( - (t2/tr)^alpha);


  pv0 = edp + pp * ( (1. - e1) * e2);
%
% Main computation loop:
%
    for j=2:1000
  t(j) = t(j - 1) + delt;
  Vv(j) = Vv(j - 1);
%
  edp = a * ( (Vv(j) - b) )^2;
  pdp = c * Vv(j) - d;
  pp = pdp/denom;
  t1 = t(j);
% Second exponential begins at t > tb
  t2 = t(j) - tb;
  if (t2 < 0.);
    t2 = 0.;
  end
  e1 = exp ( - (t1/tc) ^alpha);
  e2 = exp ( - (t2/tr) ^alpha);
  pv(j) = edp + pp * ( (1. - e1) * e2);
  end
%
plot (t,pv)
```

```
grid on
title ('Isovolumic Ventricular Pressure')
xlabel ('Time [s]')
ylabel ('Ventricular Pressure Pv [mmHg]')
```

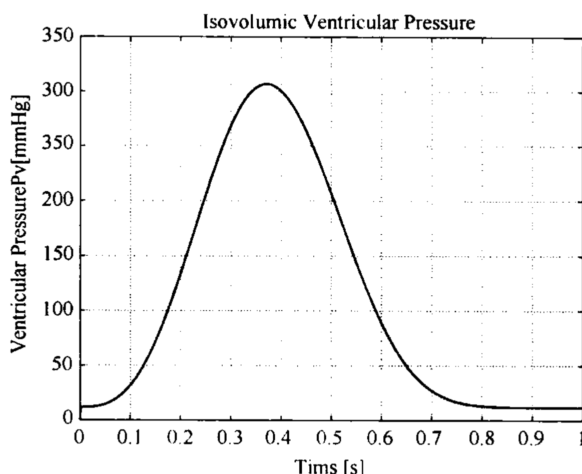


图 4.39　人心脏一个搏动周期的等容心室血压曲线 ■

4.7.4　心血管系统建模仿真

将上述左心室模型与图 4.34 所示的简化动脉血管负荷模型连接起来，就可以模拟射血过程。表 4.4 列出了正常动脉血管负荷模型的参数值。图 4.40 显示了正常狗左心室向正常动脉系统射血的过程。其中，实线（左侧纵坐标）描述了心室血压 p_v 和主动脉根部血压随时间的变化。临床中动脉血压用两个数据表示，如“110/60mmHg”分别对应于主动脉根部脉搏压的最大值和最小值，图 4.40 中这两个值约为 120/65mmHg。图中虚线（右侧纵坐标）表示心室输出流量。心室舒张末期的充盈容积为 45ml，每次搏动输出的血量（即每搏输出量）为 30ml，其射血分数（即每搏输出量与心室舒张末期容积的百分比）为 66%，对于狗这样大小的动物来说，这是正常值。

表 4.4　动脉系统模型元件的典型参数

模 型 元 件	符　　号	典 型 数 值
主动脉阻抗	Z_0	0.1mmHg・s/ml
系统动脉顺应性	C_s	1.5ml/mmHg
外周动脉阻抗	R_s	1.0mmHg・s/ml

式（4.75）表示的心室模型还可以与病变的动脉系统相连。例如，可以将动脉系统的外周阻抗 R_s 增加一倍，用于模拟动脉血管的缩窄，其计算结果如图 4.41 所

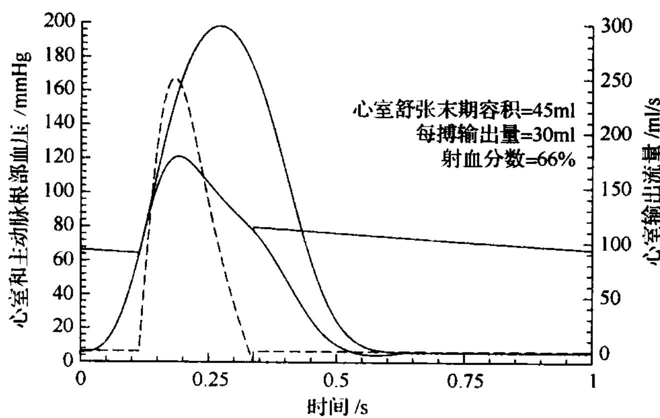


图 4.40 应用式（4.75）模型计算的正常狗左心室向正常动脉循环系统泵血时心室和动脉根部血压（实线，左纵坐标）以及心室输出流量（虚线，右纵坐标）的曲线

注：其中，峰值较高的实线曲线是动脉血管夹住时的等容曲线。心室的初始充盈容积为 45ml，每搏输出量为 30ml，射血分数为 66%，这些值都属于正常范围。

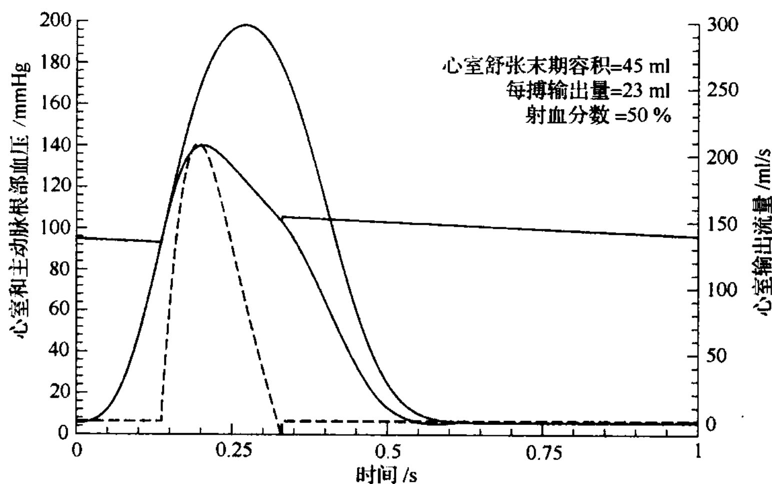


图 4.41 与图 4.40 相同的正常狗的心室模型与外周阻抗增加一倍的动脉模型相连时的计算结果

注：不出所料，反映血管缩窄的阻抗增加，导致了动脉脉搏压的增加。射血分数则从 66%降为 50%。

示。不出所料，外周阻抗的增加使动脉血压上升（图中升至 140/95mmHg），从而阻碍了心室的射血，使得射血分数降至 50%。另外，利用这个模型还可以进行动脉刚度变化等模拟试验。可见，该模型可以方便地用于阐明心脏的病理以及血管的病变。式（4.75）的心室方程由一组实验测定的参数定义，它可以描述很多种血液动力学实验现象（Palladino 等人，1997）。

虽然上述式（4.73）和式（4.74）定义的心室弹性系数的单位与式（4.72）中的定义相同，但它们的数学意义不同。式（4.75）将心室血压定义为一个解析函数，于是，根据式（4.72）的定义，其心室弹性系数 E_v（即 $\partial p_v/\partial V_v$）可以计算

如下：

$$E_v(t,V_v)=2a(V_v-b)+c\left[\frac{(1-e^{-\left(\frac{t}{\tau_c}\right)^{\alpha}})e^{-\left(\frac{t-t_b}{\tau_r}\right)^{\alpha}}}{(1-e^{-\left(\frac{t_p}{\tau_c}\right)^{\alpha}})e^{-\left(\frac{t_p-t_b}{\tau_r}\right)^{\alpha}}}\right] \tag{4.77}$$

也就是

$$E_v(t,V_v)=2a(V_v-b)+cf(t) \tag{4.78}$$

利用这个新的解析方程就可以计算各种不同心室和动脉血管状态下的时变弹性系数 E_v，包括正常和病变的心室和动脉系统，以及等容心搏和射血心搏等。图 4.42 所示就是根据式（4.77）计算得到的一组心室弹性系数曲线，图中各条曲线之间的差异相当小，并且聚集成两组；一组属于正常组，另一组则是心脏衰竭组。由此可见，这种新的弹性系数测定法可以有效地单独评价心脏的健康状况，与血管的病变分开。

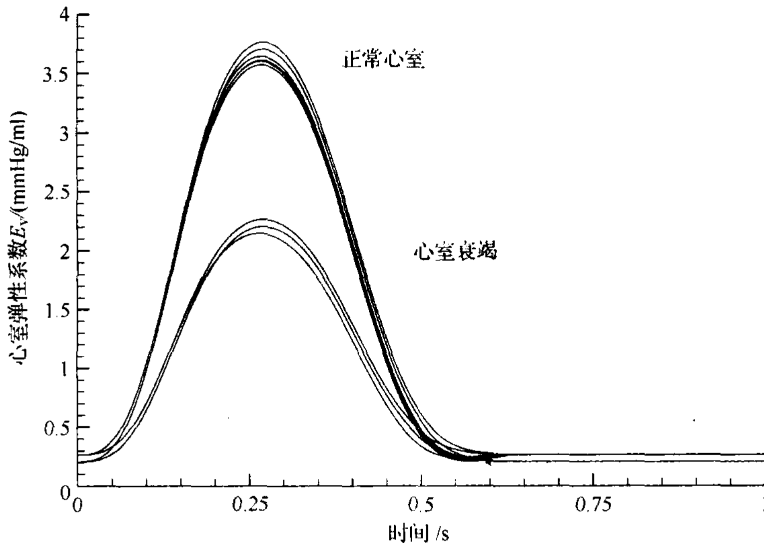


图 4.42　应用新的解析方程（4.77）计算得到的心室弹性系数曲线

注：这种方法计算的时变弹性系数可以反映心室的收缩功能，即泵血能力。

不过，实验数据表明，左心室射血时的实际心室血压仍然与模型预测值有所不同。如图 4.43 所示，射血初期实际心室压小于预测值，这被称为血压的欠激活（de-activation）；而在收缩末期心室压却较高，被称为过激活（hyperactivation）。这两个差异被称为射血效应（ejection effect）（Danielsen 等人，2000）。如果在式（4.76）的模型中加入一个流量依赖性的项就可以模拟这种射血效应，即

$$p_v(t,V_v,Q_v)=a(V_v-b)^2+(cV_v-d)F(t,Q_v) \tag{4.79}$$

这里用如下 F 函数代替了式（4.76）中的时间函数 f：

$$F(t,Q_v)=f(t)-k_1Q_v(t)+k_2Q_v^2(t-\tau),\tau=\kappa t \tag{4.80}$$

式中　k_1、k_2 和 κ——新增加的模型常数。

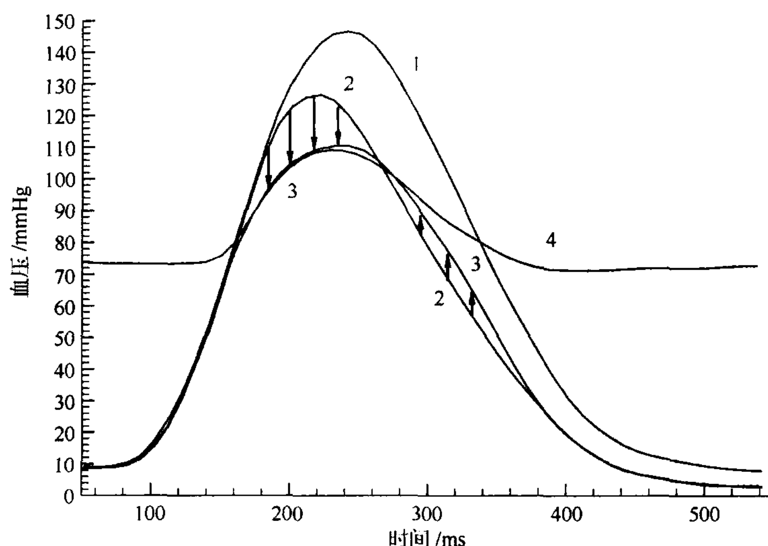


图 4.43　射血效应

注：射血初期实际心室压小于预测值，被称为欠激活；而在收缩末期实际心室压却高于预测值，被称为过激活。曲线 1 和曲线 2 分别是在初始容积和收缩末期容积下的等容收缩心室压，曲线 3 则是实际测得的射血血压曲线，曲线 4 是动脉根部血压。

总之，心室是一个压力发生器，它所产生的压力与时间、容积和流量都有关系。图 4.44 中的左图是根据式（4.79）计算得到的压力和流量曲线，而右图则是用于比较的没有考虑射血效应的曲线。可见，射血效应的加入使模型的压力和流量曲线更接近实际实验所测得的结果。

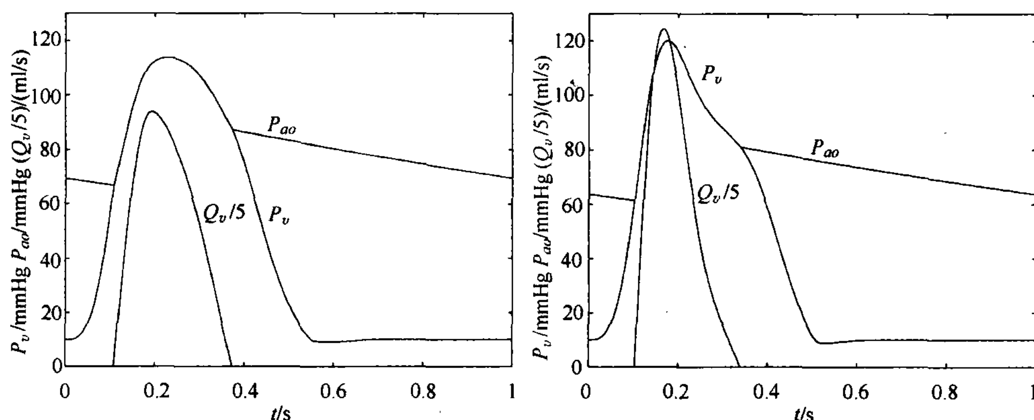


图 4.44　用射血效应修正之后的心室模型的血压和血流量曲线（左）以及相同条件下未修正的曲线（右）（为了使用与心室压相同的数值坐标，图中所示心室输出血流量为原值 Q_v 的 1/5）

如图 4.45 所示，射血效应的加入修正了心室弹性系数的计算曲线。图中两条曲线都根据式（4.77）的定义计算，虚线是没有考虑射血效应的曲线，它是对称的，

与图 4.42 相同；实线是考虑了射血效应的曲线，它不对称，向右偏斜，很像图 4.36 中 Suga 和 Sagawa 所测得的曲线。显然，射血产生的力学作用对于心脏的弹性系数具有直接的影响。

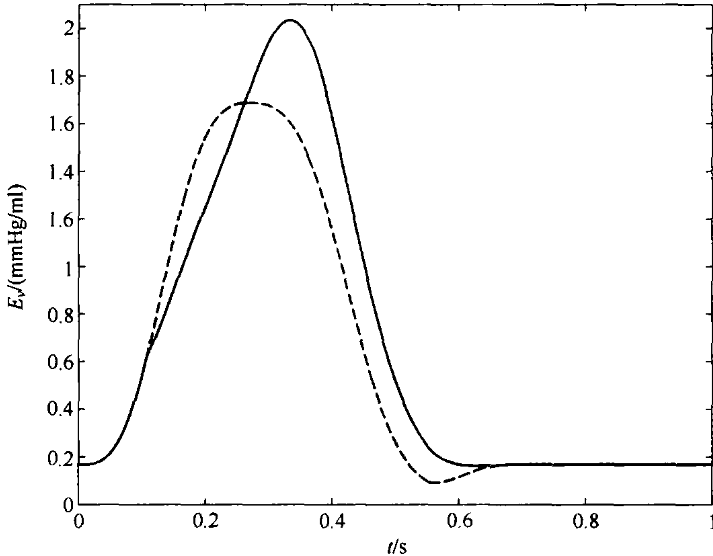


图 4.45　根据式（4.77）的定义计算的心室弹性系数曲线
（虚线是没有考虑射血效应的曲线，实线是考虑射血效应的曲线）

总而言之，左心室可以看做一个压力发生器，它产生的压力与时间、容积和流量都相关。利用实验测得的几个参数就足以描述心血管的各种动力学现象。这种建模方法将实验与理论结合在一起，形成了新的理念和新的实验。现在有人正在利用这个模型设计新的心血管健康状况测定指标。本质上，心脏射血能力的强弱应该直接与心脏的健康状况相关。

如图 4.46 所示，有人曾经利用式（4.75）的左心室模型建立了人体心脏 4 个腔室中每个腔室的模型（Palladino 等人，2000）。这个循环系统整体模型仅用几组方程和一些参数就描述了心血管生理的主要特性。例如，心室血压随心室容积变化的曲线可以表示心室所做的机械功，这种曲线称为压力-容积（Pressure-Volume，PV）环。PV 环所包围的面积就是心室所做的功。图 4.47 显示了正常心脏向正常循环系统射血时左、右心室的 PV 环（实线），显然，右心室的 PV 环比左心室小很多。图中还显示了左心室衰竭时的左、右心室 PV 环（虚线）。可见，此时左心室 PV 环明显减小，且移向容积坐标轴的右侧。由于心室衰弱时射血量减少，滞留在心脏中的血液增加，下一次搏动时心脏中的血液量随之增加，因此，图中舒张末期容积从 122ml 增加到了 194ml。根据 Starling 定律，增加心室的舒张末期容量可以增加心脏射血的压力，因此，这种心室充盈量的增加可以部分补偿心室射血量的减少。

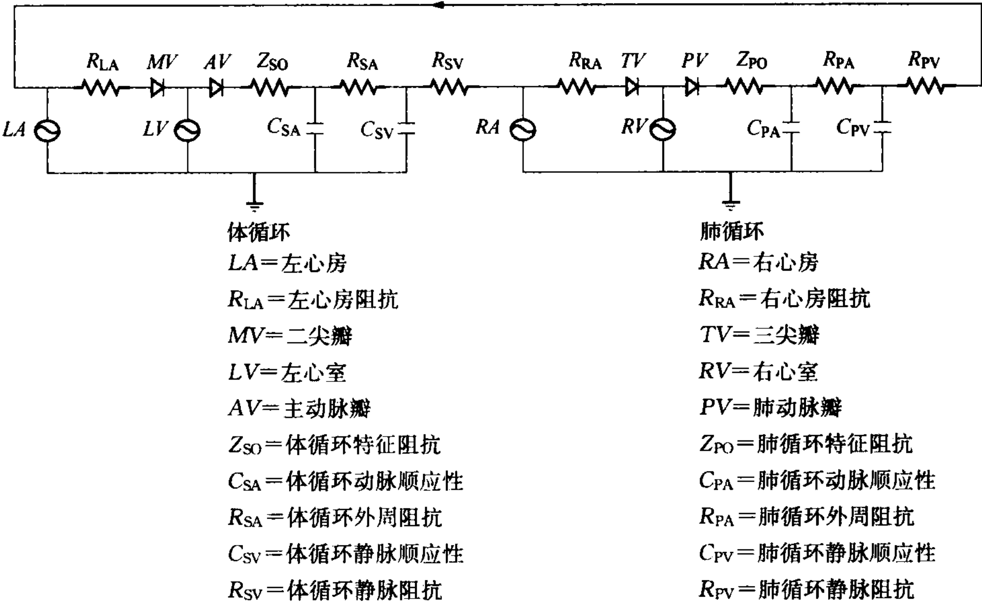


图 4.46 应用狗左心室模型描述整个人体心血管系统的血液动力学特性

（摘自 Palladino 等人，2000）

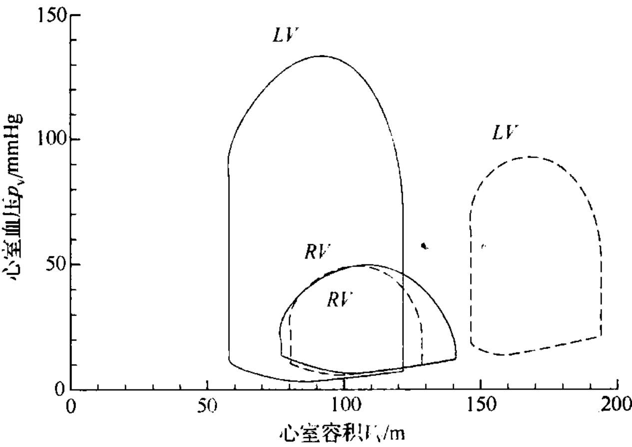


图 4.47 正常状态下左、右心室的 PV 环

（实线）和左心室衰竭时的两个心室的 PV 环（虚线）

只要改变式（4.75）的模型常数 c 的值就可以改变心室的收缩量。例如，表 4.5 列出的数据就反映了减小左心室和右心室常数 c 的值所产生的心脏充血不足。如果将左心室的收缩减小到正常值（c=1.0）的 1/3，则左心室的射血分数就会从 53%降为 25%，并且主动脉根部的脉搏压会从 133/70mmHg 降到 93/52mmHg，而左心室每搏输出量则从 64ml 减少到 48ml。根据 Starling 定律，此时左心室的舒张末期容积增加到 194ml，因此，每搏输出量得到了部分补偿，其下降量不是太大。减少左心室收缩

量与左心充血不足的效果一样，因此，肺静脉容积从1347ml增加到1981ml（表中没有显示），表明肺处于充血状态。

当右心室收缩量减为正常值的一半（即 $c=0.5$）时，也会产生类似的变化。如表4.5所示，右心室的射血分数从46%降为24%，并且肺动脉根部的脉搏压从49/15mmHg降到29/12mmHg，而右心室每搏输出量从64ml减少到44ml，右心室舒张末期容积则从141ml增加到164ml。反过来，如果使用影响肌肉收缩力的药物，也可能使心室的 c 值增大。这里没有给出详细数据，不过，利用此心脏模型可以计算心搏周期任意时刻的血压、流量和容积，它们都是时间函数。

血管特性的变化可以单独研究，也可以与心室特性的变化一起研究。另外，还可以考察心房的性能等其他血液循环系统的参数，进行其他模拟计算。此模型的模式使其可以推广应用于循环系统各个部位的详细研究。

生物力学通过数学方法将物理原理应用于生命系统，其中血液动力学研究的是人体心血管系统，该系统通过心脏的泵血功能将血液输送到人体复杂的血管网络中。在血液动力学理论的发展过程中，实验研究始终与理论分析紧密结合。实验保证了理论的正确性；而理论分析则阐明了实验现象的机理，并能够给予修正，而且还经常超前于实验研究。人类对于心血管系统的认识就在这种模式下不断进步。

表4.5 正常心脏、左心衰竭和右心衰竭状况下的心血管功能

	SV/ml		*EDV*/ml		*EF*（%）		P_{AO}/mmHg	P_{PU}/mmHg
	LV	*RV*	*LV*	*RV*	*LV*	*RV*		
正常	64	64	122	141	53	46	133/70	49/15
左心室衰竭	48	48	194	129	25	38	93/52	49/21
右心室衰竭	44	44	102	164	49	24	109/59	29/12

表中：*SV*——每搏输出量；*EDV*——舒张末期容积；*EF*——射血分数；*LV*——左心室；*RV*——右心室；P_{AO}和 P_{PU}——分别为主动脉根部和肺动脉根部的血压。注意，在各种状态下，左、右两心室的 *SV* 都相等。

习 题

1. 请写出式（4.1）～式（4.14）的MATLAB矢量表达式。

2. 请按照 z-x-y 的旋转顺序重新求取例4.2的变换公式。

3. 请分别画出图3.36所示肱骨的3个方位的受力图，对于给定的负荷和受力点位置，写出静力学平衡方程，并求解方程。

4. 设例4.5中前臂与水平线之间形成 θ 角，且 θ 从0～70°（水平线以下）以15°的增量依次变化。请编写MATLAB程序，作图表示为了维持静态平衡，肱二头肌所需产生的作用力 F_B 随 θ 的变化曲线。在此 θ 角变化范围内该作用力变化了多少？

5. 重做习题4，作图表示在同样的 θ 角变化范围内作用力 F_A、F_B 和 F_C 的曲线。

6. 分析习题 5 所得的曲线，解释为什么健身房的 Nautilus 举重机采用不对称的滑轮结构。

7. 假设图 4.9 的测力平台大小为 70cm×70cm，在踏步的某个瞬间，各个力传感器的读数分别为 $F_1=210N$，$F_2=220N$，$F_3=150N$，$F_4=180N$。请计算合力及其作用位置。

8. 根据表 4.1 的数据，计算你自己身体各肢体段相对于质心的转动惯量。

9. 重做例题 4.8，假设所用的骨钉是圆柱形钛棒，圆形横截面的直径为 9mm。

10. 请编写 Kelvin 三元件粘弹性模型的 Simulink 程序模型，并进行蠕变和应力弛豫试验，结果如图 4.23 和图 4.24 所示。

11. 应用 Kelvin 三元件粘弹性模型描述一种你所选定的生物材料的应力弛豫曲线，并利用文献提供的应力响应曲线计算模型的弹性系数 K_1、K_2 以及粘性阻尼系数 β。

12. 请编写 MATLAB 程序，求骨盆的解剖坐标系 $\{e_{pa}\}$。

13. 请编写 MATLAB 程序，利用 4.6.2 节所示的运动学数据计算体重为 25.2kg 的病人的踝关节功率。

14. 设骨盆和大腿的解剖坐标系的定义与例 4.9 和例 4.10 相同，请按照 x-y-z 的旋转顺序计算骨盆在各个方位上的角度。

15. 按照 x-y-z 的旋转顺序，重做习题 14，求解大腿的屈与伸、髋关节的内收与外展以及髋的内旋与外旋。提示：此处 $\{e_{ta}\}$ 是带 3 撇的坐标系，而 $\{e_{pa}\}$ 是不带撇的坐标系。

16. 请编写 MATLAB 程序，计算各个坐标轴方向上的 12mm 扰动对于踝关节力矩大小的影响（参见 4.6.3 节）。

17. 用 Casson 方程拟合例 4.12 的血液流变学数据，并计算血液的屈服应力 τ_0。

18. 利用表 4.4 所列的参数值，解方程（4.71），求动脉瓣关闭时的血压 $p(t)$，并作出一个心搏周期的 $p(t)$ 曲线（即 $t=0\sim1s$）。

19. 请计算初始容积分别为 $V_v=30$、40、50、60、70ml 时狗心脏的等容心室血压 $p_v(t)$，并像图 4.38 那样，将各条函数曲线重叠显示在一起。

20. 请编写 MATLAB 程序，利用式（4.77）计算心室的弹性系数，按照例 4.14 给定的参数值计算并作图显示 $E_v(t)$。

推荐阅读资料

Allard, P., Stokes, I.A.F. and Blanchi, J.P. (Eds.), (1995). *Three-Dimensional Analysis of Human Movement*. Human Kinetics, Champagne, Il.

Burstein, A.H. and Wright, T.M. (1994). *Fundamentals of Orthopaedic Biomechanics*. Williams & Wilkins, Baltimore.

Danielsen, M., Palladino, J.L. and Noordergraaf, A. (2000). The left ventricular ejection effect. In *Mathematical Modelling in Medicine* (J.T. Ottesen and M. Danielsen, Eds.). IOS Press, Amsterdam.

Davis, R.B.D. III. (1986). Musculoskeletal biomechanics: Fundamental measurements and analysis. Chapter 6 in *Biomedical Engineering and Instrumentation* (J.D. Bronzino, Ed.) PWS Engineering, Boston.

Davis R.B, Ōunpuu S, Tyburski D.J., Gage J.R. (1991). A gait analysis data collection and reduction technique. *Human Movement Sci.* 10, 575–587.

Davis R, DeLuca P. (1996). Clinical gait analysis: Current methods and future directions. In *Human Motion Analysis: Current Applications and Future Directions* (G. Harris, and P. Smith, Eds.), IEEE Press, Piscataway, NJ.

Dugas, R. (1988). *A History of Mechanics*, Dover, New York. Reprinted from a 1955 text.

Fournier, R.L. (1999). *Basic Transport Phenomena in Biomedical Engineering*. Taylor & Francis, Philadelphia.

Fung, Y.C. (1977). *A First Course in Continuum Mechanics*, 2nd Ed. Prentice-Hall, Englewood Cliffs, NJ.

Fung, Y.C. (1993). *Biomechanics: Mechanical Properties of Living Tissues*, 2nd Ed. Springer-Verlag, New York.

Gage, J.R. (1991). *Gait Analysis in Cerebral Palsy*. MacKeith, London.

Ganong, W.F. (1997). *Review of Medical Physiology*, 18th Ed. Appleton & Lange, Stamford, CT.

Greenwood, D.T. (1965). *Principles of Dynamics*. Prentice-Hall, Englewood Cliffs, NJ.

Huxley, A.F. (1957). Muscle structure and theories of contraction. *Prog. Biophys.* 7: 255–318.

Kennish, A., Yellin, E. and Frater, R. W. (1995). Dynamic stiffness profiles in the left ventricle. *J. Appl. Physiol.* 39, 565.

Mazumdar, J.N. (1992). *Biofluid Mechanics*. World Scientific, Singapore.

Meriam, J.L. and Kraige, L.G. (2002). *Engineering Mechanics*, 5th Ed., Wiley, New York.

Milnor, W.R. (1989). *Hemodynamics*, 2nd Ed. Williams and Wilkins, Baltimore.

Milnor, W.R. (1990). *Cardiovascular Physiology*. Oxford Univ. Press, New York.

Mow, V.C. and Hayes, W.C. (1999). *Basic Orthopaedic Biomechanics*, 2nd Ed. Lippincott-Raven, Philadelphia.

McCulloch, A.D. (2003). Cardiac biomechanics. In *Biomechanics Principles and Application*. (Schneck, D.J. and Bronzino, J.D., Eds.). CRC Press, Boca Raton, FL.

Mulier, J.P. (1994). Ventricular Pressure as a Function of Volume and Flow. Ph.D. dissertation, Univ. of Leuven, Belgium.

Needham, D.M. (1971). *Machina Carnis: The Biochemistry of Muscular Contraction in its Historical Development*. Cambridge Univ. Press, Cambridge.

Nichols, W.W. and O'Rourke, M.F. (1990). *McDonald's Blood Flow in Arteries: Theoretical, Experimental and Clinical Principles*, 3rd Ed. Edward Arnold, London.

Nigg, B.M. (1994). In *Biomechanics of the Musculo-Skeletal System* (Nig, B.M. and Herzog, W., Eds.). Wiley, New York.

Noordergraaf, A. (1969). Hemodynamics. In *Biological Engineering*. (Schwan, H.P., Ed.). McGraw-Hill, New York.

Noordergraaf, A. (1978). *Circulatory System Dynamics*. Academic, New York.

Palladino, J.L. and Noordergraaf, A. (1998). Muscle contraction mechanics from ultrastructural dynamics. In *Analysis and Assessment of Cardiovascular Function* (Drzewiecki, G.M. and Li, J.K.-J., Eds.). Springer-Verlag, New York.

Palladino, J.L., Mulier, J.P. and Noordergraaf, A. (1999). Closed-loop circulation model based on the Frank mechanism, *Surv. Math. Ind.* 7, 177–186.

Palladino, J.L., Ribeiro, L.C. and Noordergraaf, A. (2000). Human circulatory system model based on Frank's mechanism. In *Mathematical Modelling in Medicine* (Ottesen, J.T. and Danielsen, M., Eds.). IOS Press, Amsterdam.

Roark, R.J. (1969). *Formulas for Stress and Strain*, 6th Ed. McGraw-Hill, New York.
Singer, C. and Underwood, E.A. (1962). *A Short History of Medicine*, 2nd ed. Oxford Univ. Press, New York.
Suga, H. and Sagawa, K. (1994). Instantaneous pressure–volume relationship under various end-diastolic volume, *Circ. Res.* 35, 117–126.
Weber, E. (1846). Handwo terbuch der physiologie. Vol. 3B. R. Wagner, ed., Vieweg, Braunschweig.
Westerhof, N. and Noordergraaf, A. (1990). Arterial viscoelasticity: A generalized model, *J. Biomech.* 3, 357–379.
White, F.M. (2003). *Fluid Mechanics*, 5th Ed. McGraw-Hill, New York.
Winter, D.A. (1990). *Biomechanics and Motor Control of Human Movement*. Wiley, New York.
Vogel, S. (1992). *Vital Circuits: On Pumps, Pipes, and the Workings of Circulatory System*. Oxford Univ. Press, New York.

第 5 章　康复工程与辅助技术

Andrew Szeto 博士，专业工程师

本章目录

本章主要内容：

- 康复工程师与辅助技师在康复职业中的职责；
- 康复工程的主要研究方向；
- 残疾人的生理障碍和心理变化；
- 辅助技术评估的原则及其目标和隐患；
- 该领域的主要工程学原则和人机工程学原则；
- 该领域的职业发展机会和信息资源。

5.1　绪论

20 世纪 70 年代末以来，改善残疾人境况的科学技术得到了重大的发展，各种各样的新名称也随之出现。例如，修复与矫形、康复工程、辅助技术、辅助仪器设计、

康复技术，甚至还有生物医学工程在残障中的应用等。随着该领域的逐渐成熟，有些名词已被广泛接受，并已用于国家立法。

如今最常用的两个名称是辅助技术（assistive technology）和康复工程（rehabilitation engineering），虽然这两个名称常被互用，但两者是有区别的。该领域的先驱James Reswick（1982）曾经指出“康复工程是科学技术在改善残疾人境况中的应用”；而辅助技术则可以看作康复工程的实践产物，这种关系就像医疗保健是医学的实践产物一样。

美国公法100—407给出了辅助技术的常用定义，即“任何用于提高和改善残疾人能力的现成商品或者改装过的、定制的仪器设备和系统”都属于辅助技术。注意，这个定义将辅助技术看作有助于提高个体活动能力的各种仪器设备、政策以及服务设施，其范围很广，包括价格低廉、制作简单的低技术装置，也包括制造复杂且昂贵的高科技设备。低技术装置有双把手器具和口控操纵杆等；高技术设备则有计算机辅助的通信交流设备、人工智能阅读机、电动人造手臂等（见图5.1）。

	直接选择	扫描
无辅助	指示或手势	点头或摇头
低技术辅助	交流板	单开关交流钟
专用高技术辅助	具有语音合成和输出打印功能的辅助交流系统	具有单开关输入和语音合成输出的辅助交流系统
通用高技术辅助	具有语音合成输出的计算机	具有语音合成输出和单开关输入的计算机

图5.1 用于通信交流的辅助系统的分类（摘自Church和Glennen，1992）

康复技术、矫形术和修复术等其他几种名称也很常用。康复技术是辅助技术的一部分，是针对个体永久性或者暂时性残疾状况所导致的当前的能力限制所实施的辅助技术。传统上，矫形术用于改善肢体的功能，而修复术则是从结构和功能上用替换物代替身体的某个部分。不过，如今这两个名称广泛代表所有能够增强或者替代身体某种功能的器械设备。例如，辅助语言交流的系统有时就称为助语修复体。

5.1.1 发展历史

只要简要回顾一下历史，我们就会明白为什么辅助技术和康复工程领域会用过这么多既相似又有区别的名称。历史上，人们一直在不断利用科学技术改善残疾人的境况。第二次世界大战之后，美国人在这方面的努力目标更加明确。当时，美国退役军人管理局（Veterans Administration，VA）认识到，必须为从战场上返回的众多伤残军人做些事。然而，那时受过良好培训并且能够操作人工假肢和矫形器的技术人员实在太少，根本不能满足伤残军人的需要。为了培养这类需求量激增的技术服务人员，20 世纪 50 年代，美国联邦政府投资建立了很多修复与矫形专科学校。

美国退役军人管理局还意识到当时的人造假肢和矫形器太原始，效果不够好。20 世纪 40 年代的矫形器和假肢既笨重、不舒适，功能又很有限。于是，美国联邦政府建立了退役军人管理局假体研究委员会，其任务就是改进已有的矫形器和人造假肢装置。从此，那些以前致力于打败二战轴心国的科学家和工程师们就转向为广大民众服务，尤其为残疾的退伍军人服务。经过不懈的努力，他们开发成功了较坚固、轻巧、舒适且高效的人工假肢、电子旅行导航仪和轮椅等辅助设备。

当时美国联邦政府通过两种途径支持辅助技术和康复工程领域的发展。一是直接资助研究开发项目，利用战争期间发展的先进科学技术改善伤残退役军人的自理能力和独立生活能力；二是帮助建立修复师和矫形师培训中心，这些中心培养的人才就是如今辅助技师的先驱们。

20 世纪 60 年代初，欧洲孕妇曾经服用“反应停”（即 thalidomide）来克服早孕反应，结果导致许多畸形婴儿的出生，这一事件又一次推动了康复工程的发展。帮助严重畸形儿童恢复正常生活能力的社会需求扩大了辅助技术和康复工程的服务范围，使其服务对象除了成人之外，又增加了儿童。从此以后，医学与工程学在研究和开发中相互结合，产生了各式各样的辅助技术产品，例如，适用于不同年龄和不同性别人群的电动假肢、无臂残疾人可以驾驶的汽车以及盲人和聋人的感觉辅助器等。

1973 年，美国康复法（Rehabilitation Act）的颁布标志着政府正式将康复工程设立为一门工程学科。该法案明确授权有关部门建立了数个康复工程研究中心，这些中心隶属于国家残疾研究院，也就是后来的国家残疾与康复研究院（National Institute on Disability and Rehabilitation Research，NIDRR）。到 1976 年，美国全国大约已成立 15 个康复工程研究中心。每个研究中心都各具特色，致力于解决不同的问题，它们每年接受的研究经费资助共约 900 万美元。NIDRR 作为康复领域的主要国家政府机构，还通过康复研究与培训中心项目、研究启动基金、研究与论证项目以

及康复工程奖学金等途径（NIDRR，1999）支持康复工程和辅助技术的发展。

刚开始时，康复工程中心（Rehabilitation Engineering Centers，REC）基金用于资助大学的康复工程研究以及培养研究生高级人才。20世纪80年代中期以后，REC的资助范围拓宽到技术转化和残疾人服务。同时，退役军人管理局（VA）也针对退役军人的特殊康复需求建立了隶属其下的3个康复工程中心。VA和非VA的各个康复工程中心的研究领域包括修复学与矫形学、脊髓损伤学、上下肢体功能电刺激、盲人和聋人的感觉辅助器、组织的压力作用、康复机器人技术、技术转化、私人执照交通工具、远程通信技术、无线电技术的应用以及职业康复教育等。1979年北美康复工程协会（Rehabilitation Engineering Society of North America，RESNA）的成立又是一个里程碑，它使康复工程得到了更广泛的关注。尽管名为康复工程协会，RESNA其实是一个综合性的职业协会，它欢迎所有开发、制造、供应和使用残疾人技术的人员。因此，RESNA的会员包括作业治疗师和理疗师、医护人员、特殊教育专家、辅助技术用户等。RESNA已经成为政府的顾问机构，颁发各种标准和证书，并且，通过其举办的年会和出版的学术刊物，传播各种信息，展示最新的康复技术。随着其职责的拓展和非工程人员的加入，1995年，RESNA改名为北美康复工程与辅助技术协会（Rehabilitation Engineering and Assistive Technology Society of North America）。

尽管康复工程服务的需求很大，并且能够带来广泛的利益，但是，在20世纪80年代，这方面的服务要获得保险公司、社会服务机构和政府部门等第三方的费用支付和补偿却仍然相当困难。康复工程服务项目的补偿往往包含在那些比较容易被接受的名目下，如客户评估、假肢与矫形服务等。由于这个原因，虽然康复工程服务的需求持续增长，但当时从事这方面工作的工程师人数却并没有什么变化。

美国1986年的康复法案和1988年的辅助技术法案都特别提出了缺乏具备良好培训和经验的康复工程师的问题。这两部法案强制要求为残疾人提供康复工程服务和资金，同时，还要求为每位具有工作能力的病人安排个性化的工作和个性化的康复计划。在这两部法律之前，1973年颁布的原康复法要求合理安排残疾人的就业和中等教育，并将其定义为最少限制环境（Least Restrictive Environment，LRE）。1975年颁布的公法95－142将受教育的儿童和青少年的年龄扩大到了5～21岁，并要求为每位符合条件的残疾人制定个性化的教育计划。表5.1归纳了美国联邦政府立法中有关辅助技术和康复工程的主要法案。

表5.1 近年来美国联邦政府推动辅助技术发展的主要立法（摘自Cook和Hussey，2002）

立　法	有关辅助技术的主要内容
1973年康复法	要求联邦政府资助的企业和高等教育机构为残疾人提供合理的接纳和最小限制环境；在州政府的各种计划和为每位客户设置的个性化书面康复计划（Individualized Written Rehabilitation Plans，IWRP）中都要包括辅助技术设备及其服务；康复法第508节要求所有政府机构雇员具有使用电子办公设备的同等权利；将康复技术定义为康复工程和辅助技术设备及其服务；要求康复技术成为IWRP所包含的主要福利

（续）

立　法	有关辅助技术的主要内容
1997 年残疾人教育法修正案	每位孩子都享有自由接受合适教育的权利；残疾孩子应该与同龄人一起接受教育；给 3～21 岁受教育年龄的残疾人提供合理的接纳、最小限制环境以及辅助技术设备和服务；为每位孩子制定个性化教育计划，包括辅助技术事项；另外，还要求为 0～2 岁儿童提供服务，特别是与教育相关的辅助技术
1998 年辅助技术法（即 1998 年残疾人辅助技术法）	这是针对辅助技术设备及其服务的第一部立法；它要求消费者驱动的辅助技术服务、能力构建、倡导活动以及各州的系统设施改进；给各种辅助技术系统的发展和管理提供基金
发育残疾的辅助权利法	给各个州的发育残疾委员会、大学的相关项目以及发育残疾病人的保护提供资金；给发育残疾病人提供使用辅助技术设施的培训和技术服务
1990 年美国残疾人法（Americans with Disabilities Act，ADA）	在就业机构、州政府和地方政府机构、公共场所、商业设施、交通运输、通信等方面禁止歧视残疾人，不允许影响辅助技术的应用；某些公共机构必须配备有效的交流设施，供患有听觉、视觉或言语障碍的残疾人使用；给听觉和言语障碍的残疾人提供电话和电视设施的指示
Medicaid 医疗补助保险	这是补偿性计划；各个州政府设定的补助资格和服务范围有所不同；联邦政府设置通用的项目要求，并给州政府提供财政补助，弥补州政府的开支；成人与 0～21 岁孩子的辅助技术福利有所不同；成人的辅助技术必须包含在州立 Medicaid 计划中，或者包含在免费项目中
早期定期筛查、诊断与治疗项目	要求为 0～21 岁的人群服务；包括 Medicaid 法案中所列的必需的和可选择的服务；服务项目无需包含在州立 Medicaid 计划中
Medicare 政府医疗保险	它是辅助技术（即耐久性医疗设备）的主要资金来源；保险对象包括 65 岁以上老年人以及终身残废和完全丧失劳动力的残疾人；由联邦政府统一管理，各个州的法规一致

为了与联邦政府的立法保持一致，一些政府研究项目都努力为残疾人增加康复工程服务。例如，美国国家科学基金就曾经启动过辅助残疾人的生物工程研究项目。该项目的目标是：为残疾人提供实习设备或软件，从而提高他们的生活质量和自立能力；通过实际的设计训练提高学生的受教育水平；给大学服务于当地社区的机会。美国教育部的特殊教育与康复服务办公室曾经资助多种特殊计划和示范项目，如辅助技术评估项目、视力和听力障碍儿童的技术应用服务、交流障碍学生的交叉学科培训服务（即语言治疗师）、特殊教育以及工程教育等。1993 年，NIDRR 拨款 3860 万美元资助各个康复工程中心研究如下项目：自适应计算机及信息系统、增强型和选择型辅助交流设备、腰背痛病人的就业、听力增强辅助设备、修复学与矫形学、体能的定量测试、康复机器人技术、技术转化与技术评价、轮椅灵活性的改善、工作场所改进及调整、老年人辅助技术、残疾人私人执照交通工具、职业康复的康复技术服务、盲人和弱视病人的辅助技术以及肢残儿童的辅助技术等。在 1996 财政年度，NIDRR 共资助了 16 个康复工程中心和 45 家康复研究与培训中心，总资助金额分别为 1100 万美元和 2300 万美元。

5.1.2 信息资源

与其他新兴学科一样，早期康复工程学的知识基础都分散于各种不同的出版物。

由于其交叉学科的特性，康复工程学的学术研究论文广泛发表于《物理医学与康复文献》(Archives of Physical Medicine and Rehabilitation)、《人因学》(Human Factors)、《生物医学工程纪事》(Annals of Biomedical Engineering)、《IEEE 生物医学工程学汇刊》(Transactions on Biomedical Engineering) 和《生物力学》(Biomechanics) 等刊物。其中有些论文与实践相结合，是具体的应用；而有些论文则属于基础理论研究。20 世纪 70 年代早期，许多重要论文发表在美国退役军人管理局承办的《修复研究通报》(Bulletin of Prosthetic Research) 上，这是一本声誉很高、发行范围很广的同行评审期刊。1983 年该期刊改名为《康复研究与开发杂志》(Journal of Rehabilitation Research and Development)。1989 年，RESNA 创办了季刊《辅助技术》(Assistive Technology)，它主要发表辅助技术服务方面的文章，而不是工程师的研究性和开发性论文。为了满足有关康复工程科学研究论文发表的需求，IEEE 生物医学工程学会在 1993 年创办了《IEEE 康复工程汇刊》(IEEE Transactions on Rehabilitation Engineering)，该期刊后来改名为《IEEE 神经系统与康复工程汇刊》(IEEE Transactions on Neural Systems and Rehabilitation Engineering)，它是季刊，内容包括有关医学的康复医学、有关实际设计原理的康复技术、有关科学理论的康复科学以及神经系统。

5.1.3 康复工程的主要研究方向

康复工程领域的研究方向有多种分类方法。表 5.2 所示是根据康复工程所产生的主要辅助技术进行分类，这可能是能够反映其广度和深度的最简单方法。这些辅助技术产品的开发需要多学科研究人员的共同合作，包括机械、材料和电子等工程学科的技术人员、整形外科医生、修复师和矫形师、医护人员以及计算机专家等。例如，许多辅助设备中使用的输入和输出语音处理器由数字信号处理芯片、存储芯片以及复杂的软件构成，这些器件都是电子工程师和计算机工程师开发的。图 5.2～图 5.4显示了目前已经实现的几种辅助技术。正如本章下面几节将介绍的那样，辅助技术设备的正确设计、开发和应用需要工程师、经验丰富的临床医生、知情的终端用户（即消费者）以及护理人员的共同努力。

表 5.2 辅助设备的种类

辅助设备的种类
修复学与矫形学
人造手、腕和手臂
人造脚和腿
手夹板和上肢吊带
功能电刺激矫形器
严重视觉障碍残疾人的辅助设备
阅读与书写辅助器械（如闭路电视放大镜、电子盲文打字机、阅读器、语音计算器、替代视觉的听觉和触觉系统等）
独立行走辅助器械（如激光手杖、超声眼镜、手持式超声电筒、电子发声器、机器人导盲犬等）

（续）

严重听觉障碍残疾人的辅助设备
数字助听器
电话助听器（如 TDD 和 TTY 等）
唇读辅助器
语音一文字转换器
触觉障碍残疾人的辅助设备
缓冲垫
定制坐具
感觉替代物
压力缓冲器和报警器
增强型和选择型辅助交流设备
用户界面和键盘仿真器
专用开关、传感器和转换器
基于计算机的交流设备
语言工具和软件
操作和走动辅助器
抓取器、进食器、吊架系统、自动翻页机
环境控制装置
机器人助手
手动专用轮椅
电动轮椅、滑行车、活动躺椅
辅助驾驶装置
改装的私人执照车辆
娱乐辅助设备
手动自行车
体育运动和比赛用轮椅
改装的坐式雪橇

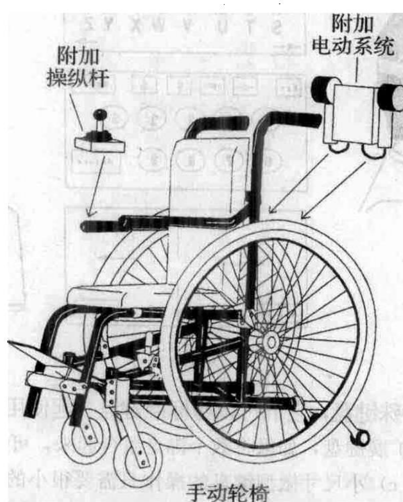


图 5.2　轮椅的附加系统（摘自 Church 和 Glennen，1992）

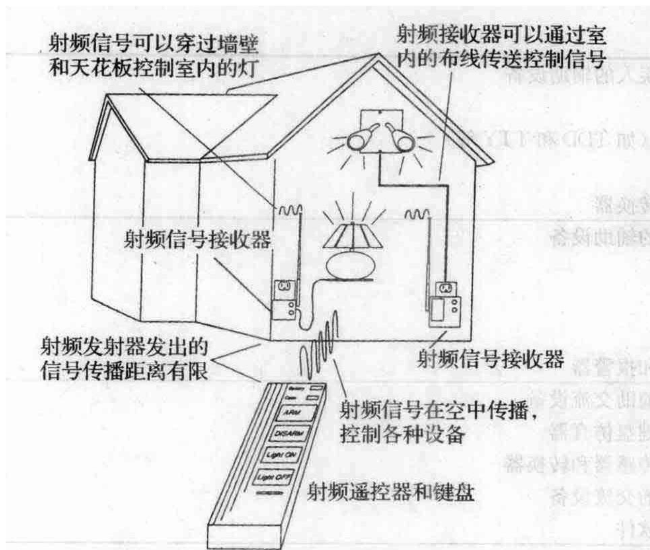


图 5.3 利用无线电射频信号控制环境（摘自 Church 和 Glennen，1992）

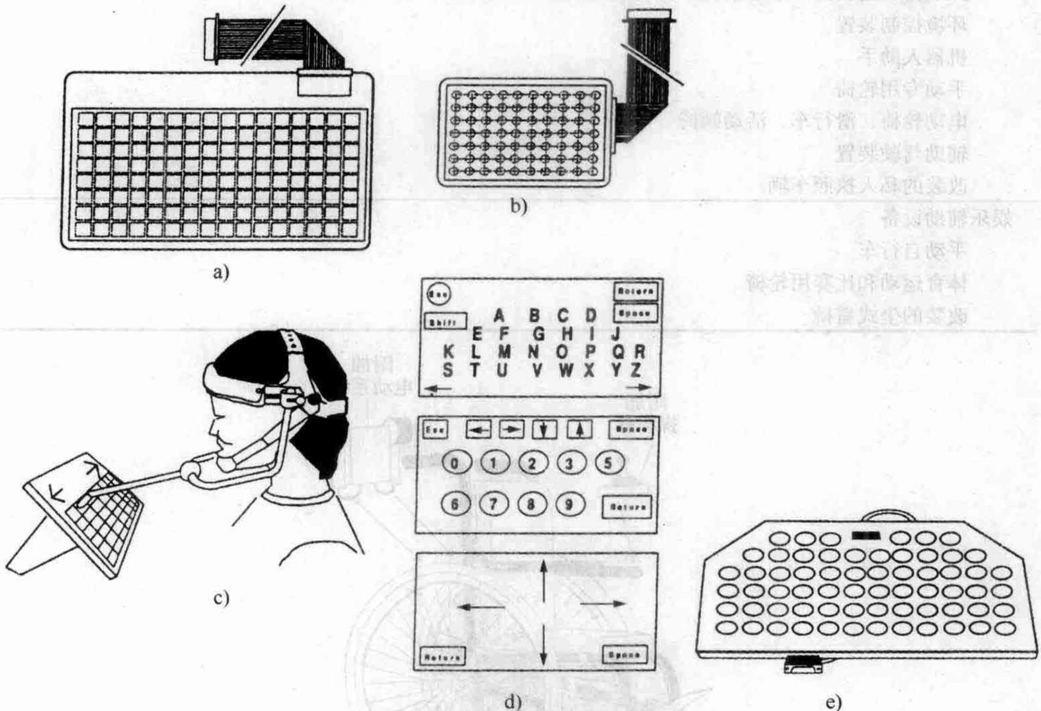


图 5.4 各种取代标准键盘的特殊键盘，它们可以与标准键盘一起使用（摘自 Church 和 Glennen，1992）
a）触摸式方键阵列组成的扩展键盘，如果将数个键盘组合起来，可以构成更大的键盘 b）紧密排列的触摸式方键组成的微型键盘 c）小尺寸微型键盘的操作只需要很小的移动范围就可以触及整个键盘 d）扩展键盘和微型键盘可以使用标准的或定制的键盘布局 e）有些特殊键盘可以直接插入计算机的键盘插孔使用，不需要特殊接口和软件

5.2 人体的要素

为了科学地应用工程学原理制作各种帮助残疾人改善困境的器械，康复工程师首先必须充分了解人体的要素以及各种损伤所带来的后果。如图 5.5 所示，人体可以看作由感受器、中央处理器和效应器组成的系统。人通过感受器接收外部环境的信息，并通过效应器来响应或者操控环境，而感受器与效应器之间是中央处理器，它具有感知、识别和运动控制功能。感知是人类解释感觉系统输入信息的过程，其产生机制包括外围神经系统中的神经回路和以往感觉经验所形成的记忆等心理因素；识别则是指与解决问题、作出决策和形成语言等相关的各种活动；运动控制是指利用上述感知和识别处理所产生的结果来形成一种控制运动的决策，然后由效应器（即神经、肌肉和关节等）来执行这种决策。效应器对外部环境产生的作用又会被感受器检测，从而在人与环境之间形成信息反馈。如果这个信息处理链上的某个部位出了问题，那么，往往就会导致残疾。表 5.3 按照残疾的解剖部位列出了各种残疾状况的普查数据。

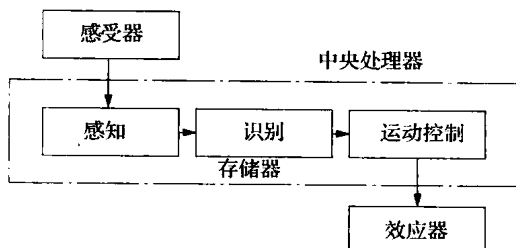


图 5.5　人体信息处理模型（其中每个模块代表可能采用的辅助技术的功能）

值得注意的是，对于人体受伤或者先天缺陷破坏了上述信息处理链上的输入感受器部分的情况，康复工程师已经有一些办法可以改善残疾状况。如果 3 种主要感觉通路（即视觉、听觉和触觉）中的某一个受到破坏，那么可以将辅助设备检测到的重要外界环境信息通过其他完好的感觉通路传送给中央处理器。例如，视觉严重受损病人使用的辅助传感器就是利用触觉或听觉将环境信息传送给使用者。这类视觉辅助传感设备包括激光手杖、超声眼镜和机器人导盲犬等。如图 5.6 所示，康复工程师在修复和改善上述人体效应器的运动障碍方面也已经取得了成功，这类辅助设备有人工上肢和下肢、各种轮椅、环境控制器以及将来可能实现的机器人助手等（见表 5.2）。

表 5.3　美国残疾人残疾状况普查数据
（数据取自 Stolov 和 Clowers，1981）

45～50 百万人轻度残障
32%听力
21%视力
18%脊髓
16%腿部和髋部
5%手臂和肩部
4%言语
3%瘫痪
1%截肢
7～11 百万人严重残障
30%脊髓
26%腿部和髋部
13%瘫痪
9%听力
8%视力
7%手臂和肩部
4%截肢
3%言语

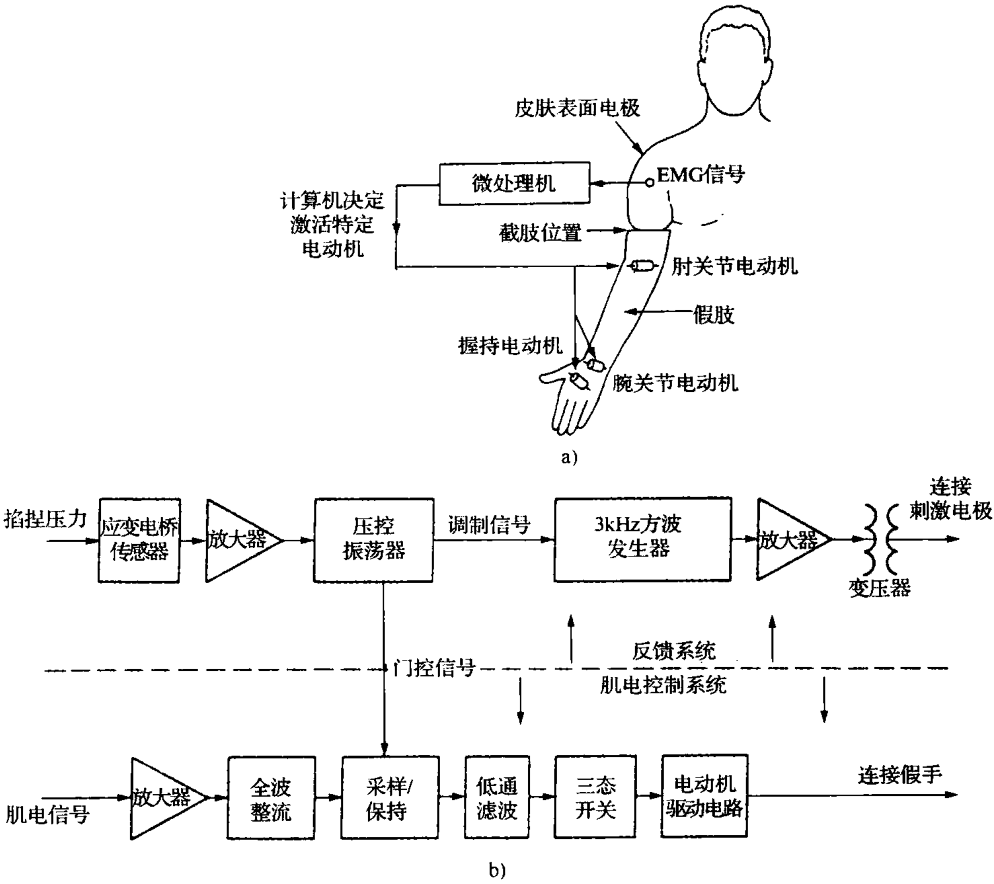


图 5.6 图 a 的系统能够从一组肌电电极记录信号中提取多种时间信息用于控制多个电动机；图 b 为一种可以提供作用力反馈信号的前臂电刺激系统（摘自 Webster 等，1985）

不过，如果残疾部位在人体的“高级信息处理中心”，即上述中央处理器部位，那么，辅助技术至今仍然收效甚微。例如，病人中风导致了失语症，康复工程师和语言治疗师就无能为力，不能恢复其有效的交流能力。尽管已经开发了各种各样的新型高级辅助交流设备，但是没有一种能够替代人的自我意识。如果使用者自己没有能力形成语言信息，那么辅助交流设备就毫无用处。

此外，康复工程和辅助技术的目标是缓减残疾所造成的后果，因此，康复工程师必须认识到心理调节对于慢性残疾人康复的重要性，必须了解残疾人的心情和心理状态，这样才能使所提供的或推荐的辅助方法适当、及时、可行，并最终可以被采纳。

慢性伤残病人受到的最大打击就是被社会孤立，致使其社会地位下降。这种社会地位的丧失可能来自残疾的直接影响（即社会隔离）或者间接影响（即经济挫折）。因此，残疾病人除了承受个人收入的大幅下降以外，还必须与残疾引起的 3 种主要心理状态进行斗争，那就是丧失自尊、过于依赖他人和消极情绪。

因受伤而致残的病人，其心理调节过程一般要经历如下 5 个阶段：震惊、认识、否认、承认和接受。在伤残发生的开始几天，病人通常很震惊，对外界反应冷淡，对于所发生的事情没什么了解。此时与病人谈论康复事宜一般没什么效果。

数周或数月之后，病人会开始逐渐认识到残酷的现实，并表现出忧虑、恐惧、甚至惊慌失措等情绪反应，还可能偶尔表现出沮丧和愤怒。在这种情感状态下，如果急切地进行康复干预也不会有什么作用。

下一阶段，为了在心理上不被忧虑和恐惧击垮，病人会采取逃避的态度，主要是否认事实，宣称伤残只是暂时的，会完全恢复。这种否认的态度会在伤残发生之后持续很长一段时间，并且反复发生。

等到病人确切地了解了伤残的性质及其限制和可能的后果之后，就会承认事实。在这个时期，虽然病人完全了解了自己的伤残状况，但可能还不能接受。对现实的逐步认识经常会伴随着沮丧情绪，结果使病人对原来喜欢的很多事物都失去兴趣。

最后，残疾病人的心理状态会调整到适应期，也就是接受期。此时，病人已经经历了伤残之后的各种主要情绪反应，认识到伤残可能带来的限制，并且在心理上已准备要发挥自己最大的潜能。在这种情况下，康复工程师和辅助技师的干预就比较合适且有效。如果能够让残疾病人参与需求的确定、计划的制定和方案的选择等工作，那么不论在心理上还是生理上，对病人都非常有利。

5.3　辅助技术评估的原则

康复工程师不仅需要知道辅助设备的物理设计原理，而且还必须遵守残疾人辅助技术应用的主要原则。也就是为了使辅助技术的应用取得成功，技术人员必须仔细考虑病人的需求、爱好、能力、限制条件，甚至病人的周围环境等因素。在辅助技术应用中至少存在如下 5 个主要误解：

误解 1：辅助技术可以解决所有问题。 虽然辅助设备可以使某些操作变得比较容易完成，但是仅仅依靠辅助技术并不能减轻残疾带来的所有困难。

误解 2：患有同样伤残的病人需要使用相同的辅助设备。 由于致残相同的病人可能具有完全不同的需求和爱好，辅助技术的应用必须个性化（Wessels 等，2003）。

误解 3：辅助技术必定是复杂且昂贵的。 有时低技术含量的设备反而是最合适的，甚至会因为它们的简单、便于使用和维护且成本低而更受欢迎。

误解 4：辅助技术的设置总是精确且最佳的。 许多经验明确表明残疾人的辅助技术设备是不精确的，需要随时调整。根据使用者的健康状况、生活环境、爱好等状况，使用者和康复专家需要定期评估辅助设备的使用情况，随时作出相应的调整（Philips 和 Zhao，1993）。

误解 5：任何残疾状况都要使用辅助技术。 根据 1990 年美国人口普查局的国民健康问卷调查数据，在非维持生命必需的辅助设备中大约有 1/3 没有被使用，或者开始被使用 3 个月之后即停用了。

除了避免上述误解之外，康复工程技术人员还应该遵守如下有助于病人与辅助技术的恰当匹配的基本原则，坚持这些原则能够提高病人接受并充分利用辅助技术的几率。

原则1：在辅助技术干预过程中，必须尽早明确地定义、罗列并整合使用者的目标、需求和操作任务。 为了避免忽略使用者的各项需求和目标，应该使用预先制定的表格清单。本章末尾所列的推荐阅读资料中就有一些有用的评估表。

原则2：具备不同技能并了解操作方法的专家的共同参与，可以增加辅助技术应用取得成功的可能性。 根据辅助设备的使用目标和环境条件，应该有一些专家参与辅助技术与病人需求之间的匹配处理工作。表5.4列出了各种辅助技术及其涉及的专业。

原则3：必须全面评估和分析使用者的爱好、智力和体力上的优势和弱势、居住条件、对辅助技术的忍耐性以及未来可能的变化等因素。 康复工程师会发现那些非技术人员用定性语言描述的内容必须转化为能够测量和定量计算的特性参数。例如，只有根据残疾人上肢的伸展、抬举和握持能力的定量计算结果，才能确定他是否可以使用人工上肢。

表5.4 辅助技术的研究领域及其涉及的专业

辅助技术研究领域	负责人员的专业①
学术技能和职业技能	特殊教育 职业康复 心理学
辅助交流技术	言语与语言病理学 特殊教育
计算机的使用	计算机技术 职业康复
日常生活技能	作业治疗 康复技术
特殊适应性	康复工程 计算机技术 修复术和矫形术
走动	作业治疗 物理治疗
坐与立	作业治疗 物理治疗
书面交流	言语与语言病理学 特殊教育

① 根据所遇到技术问题的复杂程度，可以在专业负责人员中添加辅助技师或者康复工程师。

原则4：为了避免遗漏可能的有效解决方案，必须全面仔细地考察满足用户要求的可用辅助技术。 电子数据库，如各种辅助技术的网站和主要技术供应商的网站等，往往能够给康复工程师和辅助技师提供可能采用的设备的概况，经过改造，可以用于消费者。

原则5：选用辅助技术设备时，必须考虑使用者的爱好和选择。 调查结果表明，辅助技术被拒绝使用或使用不当的主要原因是对使用者的需求和爱好考虑得不够周全。在寻求合适的技术手段的过程中，应该将技术的最终使用者看作合作伙伴和利益相关者，而不是将他们看作被动的、无所谓的服务对象。

原则6：辅助技术设备必须在以后实际使用的主要场所和环境中定制和安装。 使用场合中看似细小、无大碍的事物往往可以决定辅助技术应用的成败。

原则7：不仅要训练使用者能够使用辅助设备，而且还必须使护理人员和家属了解设备的用途和限制条件。 例如，辅助交流设备通常要求参与交流的人采取一种

不同的交流方式，改变他们通常的习惯，以便让使用设备的残疾人能够表达更多的思想，甚至在交流过程中能够扮演主动的角色，比如引入新的话题或者改变话题。除非护理人员或者家属改变他们的交流方式，否则无论这种辅助交流设备多么有用，刚使用设备的残疾人都会受到劝阻。

原则 8：必须定期回访，重新调整设备，并评估用户的使用情况。 刚开始使用辅助技术设备的半年时间里，使用者及其周围其他人需要适应新设备。如果相关人员和环境发生了变动，辅助设备原来的工作状态就可能不再合适，设备就需要重新调整和优化。定期回访和调整可以减小辅助技术被放弃的几率，从而减少时间和资源的浪费。

5.4　康复工程设计原则

为了缓减各种各样的伤残状况，在设计辅助技术设备时，必须综合利用不同学科的知识和技术。康复工程学的这种内在跨学科性质使这个新兴领域很难具备普遍适用的基本原理。最相关的原理往往取决于特定的具体问题。例如，在设计电动轮椅的环境控制系统时，最重要的就是电子和通信工程方面的知识和原理；但是在为脊髓损伤导致上肢功能丧失的病人开发植入式功能电刺激矫形器时，最有用的就是神经肌肉生理学、生物力学、生物材料以及控制系统等方面的知识和原理。

无论面对怎样的残疾状况，康复工程的目标都是设计辅助设备，也就是先确定用户的需求，再设计辅助设备用以满足这些需求，这是一种创作过程。系统性的创作方法对于成功完成康复工程项目非常重要，设计过程主要包括如下 5 个关键步骤：分析、综合、评估、决策和实施。

分析

缺乏经验但满怀热情的康复工程专业学生面对残疾人的求助和要求，往往会立即思考可能的解决办法，而忽略最重要的第一步工作，那就是仔细分析问题和需求。结果，在经过许多徒劳之后，他们才明白，必须先全面调查所需解决的问题，才能找到有意义的解决方案。康复工程师必须首先确定辅助设备使用的场合、时间以及使用的频繁程度。使用环境怎样？其他人是如何解决类似问题的？环境约束条件（如大小、速度、重量、部位、物理界面等）有哪些？心理方面的制约条件（如使用者的爱好、他人的支持、忍耐能力、认知能力等）又有哪些？经济上如何考虑？商品价格、租金、试用期费用怎样？维护和维修又如何安排？要回答这些问题，就必须仔细研究并计算相关的数据。例如需要提举的物体的重量和大小，操作对象的形状和质地，以及理想设备所需达到的操作性能等。优秀的分析工作最终应该列出“理想”解决方案所具备的操作性能和技术指标。在随后的各个设计阶段，这份性能指标清单就是选择最佳解决方案的重要指南。

例 5.1　请为一位轮椅用户设计一个电动机械装置，可用于上下升降其下肢，以防下肢产生水肿。请列出该装置的性能指标。

解：以下是该机械装置性能指标的一种写法：

- 能够在 5s 内上下活动一次下肢；
- 可以由坐在轮椅上的人独立操作；
- 有一个紧急停止开关；
- 要与现有的轮椅及其脚蹬兼容；
- 工作时无噪声；
- 总重量不超过 5lb⊖。 ■

综合

康复工程技术人员如果能够阐明问题的本质，那么他对问题的解决方法很有可能就有了思路。虽然不是严格的前后顺序关系，但是，分析问题之后通常就要对各种可能的解决方案进行综合。这种综合是一项创造性很强的工作，需要遵守前面所述的各项工程设计与应用的基本原则，并且需要参照各种相关手册、设计刊物和产品目录，并咨询其他专家。在制定和评估各种可能的解决方案的过程中，随着对问题逐步深入的了解，设计人员往往会不断想出以前没有考虑到的新方法。建议将这种综合分析过程中获得的每一种待试验方案绘制成草图，并给出详细的技术描述。

评估

根据问题的复杂性以及时间、资金等条件，对 2～3 个最有希望的方案应该进行进一步评估，评估方法可以采用仿制品的现场试用、计算机仿真或者详细的机械图样等。在整个评估过程中，设计者应该与最终用户和其他利益相关者进行磋商。现场试用的实验结果应该用录像等方式详细记录下来，便于事后重新查看。一种很有效的方案评估方法是作出定量比较图，比较各种方案满足甚至超过原定性能指标和操作特性的情况。

决策

最终的方案一般总是有所妥协，如果明白这一点，那么最终方案的选择就会比较容易。比较了各种有希望的解决方案之后，可能会有几种方案看起来满意程度相当。在这种情况下，最终的决策就要根据使用者的爱好或者其他难以预料的因素决定。有时可以咨询遇到过同样问题的其他人。不过，最重要的是要慎重考虑使用者的爱好，这就是辅助技术应用原则中的第 5 条。

实施

根据最终选定的方案制作、匹配并安装辅助设备时，需要另外制定项目计划。根据项目的规模，这种计划可以是一张简单的任务清单，也可能是一套复杂的规划，较大的规划往往需要很多具备不同技能的专家的参与才能实现。

例 5.2 请列出制作例 5.1 所述的电动下肢升降器时所需的主要设计步骤。

解：主要设计步骤如下：

- 用于提升轮椅脚蹬的机械连接装置的设计；

⊖ 1lb=0.45359237kg。

- 根据静力学原理计算抬动使用者下肢所需的力；
- 计算电动机所需的传动系数和扭矩；
- 计算轮椅所需配备电池的功率；
- 购买机电零件；
- 制作机械部件和电子部件；
- 装配、调试，或者重新设计；
- 现场试用并评估样机。 ■

5.4.1　工程学主要原则

康复工程学涉及的每门学科及学科分支都有自己的一套重要设计原则，在项目设计时必须予以重视。例如，开始设计数字电路时，必须首先确定使用哪种逻辑器件、需要同步还是异步时序电路等问题。有些常规硬件设计事项在各种设计任务中都需考虑，如最坏情况设计、计算机仿真，温度效应、可靠性、安全性等。其中，最坏情况设计就是指各种器件的性能下降时，电子系统或者机械系统的运行仍然必须保持足够的稳定性。计算机仿真和计算机辅助设计软件 CAD 常用于预测各种不同性能和尺寸的器件组合形成的整个电子系统的运行情况。

同时，在设计时还必须考虑温度和环境因素对于系统性能和可靠性的影响。例如，在极端的温度条件下，电池的容量会降低。温度也可能影响系统的可靠性，因此，应该适当采用通风散热器件，以防止温度过高。此外，为了提高系统的可靠性和耐用性，在最终的设计中应尽可能消除导线和接插件中的应变效应。

产品的安全性也是一项非常重要的设计原则，尤其是对于康复工程和辅助技术。机电系统必须安装应急开关，发生紧急情况时可以立刻中止设备的运行。要用熔丝和大规格配线，以便留有余量，保证安全。另外，要使用机械制动器和联动装置，确保各个部件连接正确，防止机械部件出现危险的或不适当的活动。

如果所设计的辅助设备需要抬举或者支撑身体的部分重量，那么就必须进行生物力学的静力学和动力学分析。最简单的分析就是计算在静止稳定状态下支撑物体或人体所需的静态作用力，这种基本工程分析一般包括如下 3 个步骤：计算作用在物体或人体某部位的作用力矢量；计算力臂；确定各个部件和人体肢体段的重心位置。静态时，所有作用力之和以及力矩之和都等于零。动态时，系统的控制方程是牛顿第二运动定律，也就是作用力的矢量和等于质量乘以加速度矢量（即 $\boldsymbol{F}=m\boldsymbol{a}$）。

例 5.3　如图 5.7 所示，假设某人体重 125lb，仰卧在一块平板上，平板架在两个间距为 72in 的支条上。如果人体下肢的重心位于下肢中心线膝盖骨以上 1.5in 处，请根据图中所示数据计算此人右腿的重量。

解： 当此人双腿平放在平板上和右腿抬至垂直于平板位置时，分别记录秤的读数 S_1 和 S_2。以左边支条为转轴求力矩之和，可得静力学方程

$$WD=L\ (S_1-S_2)$$

式中　W——右下肢的重量；

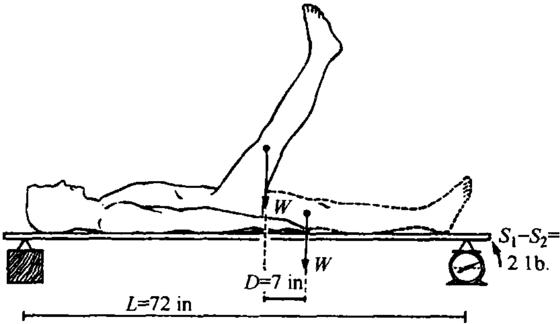


图 5.7　用平板和秤测量人体各个肢体段重量的方法

L——平板两支撑条之间的距离；

S_1——两腿都平放在平板上时秤的读数；

S_2——右腿抬起时秤的读数；

D——右腿抬起时下肢重心的水平移动距离。

假如秤的两次读数分别为 $S_1=58\text{lb}$，$S_2=56\text{lb}$，且 $D=7\text{in}$；将这些数据代入以上方程，就可以求得右腿的重量为 20.6lb。■

例 5.4　如图 5.8 所示，假设某病人利用墙上安装的滑轮装置锻炼肩关节伸肌。载重盘重为 4lb，上面加了重量为 20lb、10lb 和 5lb 的砝码，病人施加的作用力为 45lb。请计算整个系统总的合力，以及载重底盘加速度的大小和方向。

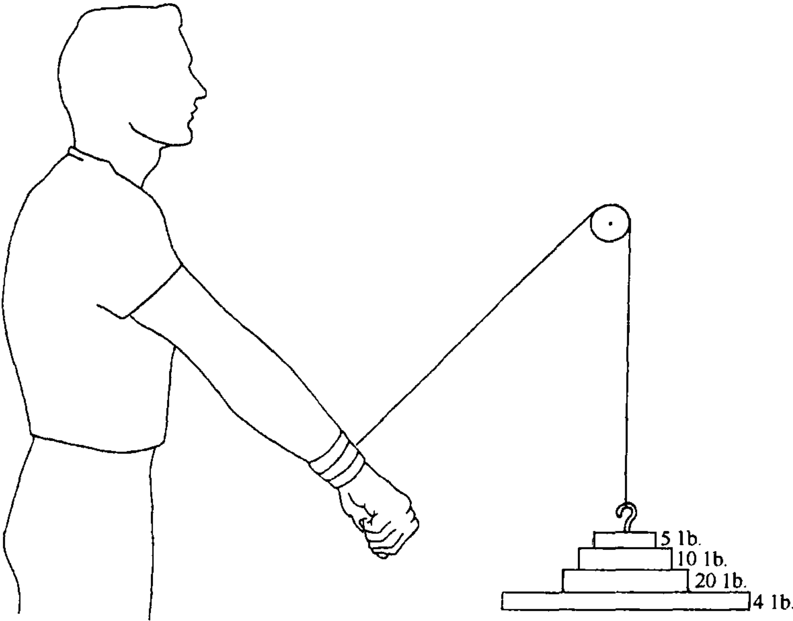


图 5.8　病人利用墙上安装的滑轮装置锻炼肩关节伸肌（摘自 Le Veau，1976）

解：所有砝码和底盘的作用力之和 39lb 的方向垂直向下，而滑轮绳索上 45lb 拉力的方向则是垂直向上，于是，净合力就是 6lb，其方向垂直向上。根据牛顿第二运动定律 $F=ma$，其中，m 是砝码和底盘的质量之和，a 是砝码和底盘的加速度。将重量 39lb 除以重力加速度 $32.2\mathrm{ft/s^2}$㊀，可以求得质量 m 为 1.21slug㊁。将这些数值代入方程 $a=F/m$，就可以求得底盘向上的加速度为 $4.96\mathrm{ft/s^2}$。■

5.4.2 人机工程学主要原则

人机工程学（Ergonomics），即人因学（human factors），是康复工程和辅助技术设计中另一个不可缺少的部分。在进行辅助技术设计时，尤其是设计工具、调节器、电子器械、作业和界面时必须考虑人的行为、能力、限制条件等因素。残疾人比正常人更难以操作设计欠佳的辅助器械。下面介绍一些与康复工程关系特别密切的人机工程学原则。

正确体位原则

如果体位或者支撑不正确，那么，丧失保持稳定姿势能力的残疾人，在重力作用下，其形体畸变和功能限制会加重。例如，如果手臂丧失正确的支持力，那么，即使操作放大了的键盘都会很慢且易出错。人体躯干上部稳定性的丧失也会限制手臂的功能，因为躯干对人体具有支撑作用。

在辅助技术的每个设计阶段，不论计划使用何种辅助技术，康复工程师都必须确保人体的躯干、腰部、后背、腿和手臂始终具有必要的支撑和稳定（见图 5.9）。如果体位的支撑是个问题，那么，在设计初期就应该咨询理疗师或者熟悉病人情况的职业治疗人员。表 5.5 列出了需要考虑坐姿和体位的常见残疾人状况。

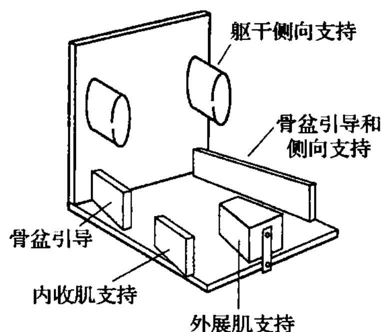


图 5.9　正确体位的坐具调整器（摘自 Church 和 Glennen，1992）

表 5.5　需要考虑坐姿和体位的残疾人状况（摘自 Galvin 和 Scherer（1996）第 66 页）

病人状况	特　征	坐姿问题
脑性瘫痪	非进行性神经肌肉损伤	
肌张力过高型	固定性畸变、活动减弱、畸形	矫正畸形、正直体位、降低肌张力
肌张力过低型	半脱位、活动减弱、关节活动过度	提供正直体位的支撑，增强肌肉的控制力
手足徐动型	活动过度、稳定性减弱	提供稳定支撑，但允许可控的功能性活动

㊀ $1\mathrm{ft/s^2}=0.3048\mathrm{m/s^2}$。

㊁ slug 是质量单位，它表示在 1lb 重的力作用下产生 $1\mathrm{ft/s^2}$ 的加速度时的质量。——译者注

（续）

病人状况	特　征	坐姿问题
肌营养不良症	进行性神经肌肉疾病	
Duchenne 型肌营养不良症	从近端至远端丧失肌肉控制	提供稳定的坐具，使病人能够找到平衡点
多发性硬化症	恶化与缓解反复发生	设计较灵活的系统以适应需求的变化
脊柱裂	先天畸形，一处或多处脊椎弓缺失，造成感觉减弱或丧失	减少压迫风险，允许完好的上肢和头部的控制
脊髓损伤	脊髓受损，受伤部位以下的功能完全或者部分丧失，稳定之后属于非进行性损伤，感觉减弱或丧失，可能发生脊柱侧凸或者脊柱后凸	减少压迫风险，允许躯干的功能性活动
成骨不全症	骨脆病、结缔组织疾病、功能范围受到限制、多重骨折	提供保护
外科损伤	固定或者定形	用支架固定或者用矫形器定形
创伤性脑损伤	严重程度取决于中枢神经系统的损伤，可能导致认知障碍，稳定之后是非进行性损伤	功能改善性康复过程，建立灵活的系统，以适应需求的变化
老年人普通人老龄化	常见固定性脊柱后凸，骨质疏松，体力减弱，大小便失禁	提供舒适的支持物，保持水分，适应脊柱后凸
残疾人老龄化	例如，患有脑瘫的老年病人可能具有固定性畸变	提供舒适的支持物，支撑畸变

解剖控制部位原则

由于辅助设备必须在使用者的指令控制下才能运行，因此，使用者必须能够通过明确的、意志控制下的动作来可靠地表明自己的意愿。现在，利用已有的各种各样的开关和传感器，可以选用残疾人身体上能够给出可靠的控制信息的任何部位作为辅助技术的解剖控制部位。一旦最佳部位选定之后，利用各种传感器、开关、操纵杆和键盘等就可以设计合适的界面接口。除了手指、肘、肩膀和膝盖等比较明显的控制部位以外，有些细微的动作，如扬眉、绷紧某块特定的肌肉等，也都可以作为辅助设备的控制信号。一般地，应该对多个有希望采纳的控制部位进行定量分析，比较其控制动作的相对速度、可靠性、特异性以及重复性。如果用仿制模型、秒表、卷尺和摄像机现场测试一下，那么对于采集各种性能数据就会很有帮助。

假如残疾人的身体条件无法做到从一组给定的选项中直接选择某一项，那么，通常可以用某个解剖部位控制的单开关激活方式与自动扫描相结合的方法来实现选择。其中一种扫描方式是行-列扫描，也就是各个选项排列成一个阵列，按照从上至下的顺序逐行依次闪亮，当包含所需选项的行闪亮时，使用者按一下开关选中该行；然后该行所包含的选项再从左至右闪亮，直到所需选项闪亮时，第二次按下开关即可选定。例如，二维字母阵列的行一列扫描选择方式可以用于编辑语句，其编辑速

度与字母在阵列中的排列位置有关。以下例 5.5 介绍了两种常用的字母排列方式，即按字母顺序排列的方式和按字母出现频率排列的方式。

例 5.5　如下表所示，有一种辅助交流设备既可以按字母顺序排列的方式扫描，也可以按字母出现频率排列的方式扫描，其行—列扫描方法如下：必须按动开关两次才能选中阵列中的某个字母；扫描从顶行开始，每行闪亮的停留时间是 1.5s；行扫描从左边开始，每项的停留时间也是 1.5s；每完成一个字母选择之后，随即从顶行开始重新扫描。假设要编辑"I WANT TO GO TO SEA WORLD"这个句子，请分别计算无出错的情况下上述两种字母排列方式所需的扫描时间。

按字母顺序排列的扫描阵列

空格	A	B	C	D	E	F
G	H	I	J	K	L	M
N	O	P	Q	R	S	T
U	V	W	X	Y	Z	TH
IN	ER	RE	AN	HE	.	,

按字母出现频率排列的扫描阵列

空格	E	A	I	L	HE	Y
T	O	S	D	P	AN	ER
N	R	C	F	IN	ES	Q
H	TH	M	B	V	X	Z
U	W	G	K	J	.	,

解： 编辑题中给定的句子所需的时间与选择句子中每个字母和空格所需的扫描停留步数成正比。选择字母"I"时，按字母顺序排列的扫描共需要 5 步（即第 2 行第 3 列）；按出现频率排序的扫描也需要 5 步（即第 1 行第 4 列）。插入空格时，两种方式都需要 2 步（即第 1 行第 1 列）扫描。选择字母"W"时，两种方式的扫描步数也相同，都是 7 步。而对于字母"T"，按字母顺序排列的扫描需要 10 步，而按出现频率排序的扫描只需要 3 步，因此，每输入一个"T"，按频率排序的扫描方式要节省 7 步，即 10.5s 时间。这样，在不出错的情况下，编辑题中所示句子时，按字母顺序扫描方式共需 213s 时间，而按频率排序扫描方式只需 180s 时间。由此可见，即使对于只包含 7 个单词的短句，用频率排序方式也可以节省 30s 时间，如果再利用"AN"等双字母组合代替单字母输入，那么就可以节省更多的时间。 ■

简单直观操作原则

仪器设备设计的一个通用原则就是操作要简单直观，基于电子器件和计算机的辅助技术设备更应如此。简单直观的关键在于设备控制和显示的兼容性和最优化。兼容性是指各种控制动作和指示分别与设备响应的期望结果相一致的程度。如果辅助设备具有较好的兼容性，那么，设备的使用者就能够学得快、反应快、出错少，从而提高用户的满意度。虽然人们确实可以适应那些与他们的习惯不一致的事物，但这是要付出代价的，例如，出错会增加，工作速度会变慢，并且操作时需要更集

中注意力。因此，康复工程师必须了解并遵循常规兼容性关系和人机工程学的基本指导方针，例如：

- 显示及其相应的控制应该具有一定的物理相似性。
- 显示及其相应的控制应该具有相似的物理安排并具有指示和标记。
- 显示及其相应的控制的移动方向应该一致，且应该位于同一个平面上，如转盘应该与旋转式显示器相匹配，而垂直滑杆应该与垂直显示器匹配。
- 开关或转盘的相对动作应该与人们日常的操作习惯相一致，如向上表示“开”、顺时针旋转表示增加、刻度从左至右是增加的方向等。

表 5.6 列出了选择各种显示器的其他指导原则。

表 5.6 选择视觉显示器的基本指导原则

（摘自 Sanders 和 McCormick 的《工程与设计中的人因学》第 7 版，1993）

显示内容	显示方式选择	理由	举例
允许、禁止；启动、停止；开、关	灯	可以清楚地表达开或关的状态	
识别标志	灯	容易看见，可以用间隔、颜色、位置、闪烁频率等进行编码，在仪器面板上也可以用标签指示	
提示或报警	灯	引人注目，如果亮度足够，远处也能看见；为了醒目，可以用闪烁方式显示	
文字指示（如操作顺序等）	灯式指示牌	简单的动作指示，可以节省决策时间	RELEASE EJECT
准确的数量	数字计数器	只显示一个数值，因此，减小了误读的几率	5 2 9 0 0
大概数量	固定刻度盘上的指针移动	从指针的位置可以快速获得大概的数量信息及其相对变化率信息	0 10 20 30 40 50 60 70 80 90 100
分档量化	固定刻度盘上的指针移动	控制与显示动作之间具有自然的关系	4 5 6 7
跟踪显示	固定标志上的单点光标或者十字光标	提供误差信息，易于纠正	

（续）

显示内容	显示方式选择	理　由	举　例
飞行器姿态	飞行器相对参照物的机械式或电子式位置显示	直接比较自身位置与已知参照物或者基线之间的关系	DOWN

指示器的适宜性原则

在设计传递信息的指示器时，人体感觉通道的选择有时已有定论，比如为视障病人设计报警指示时就是如此。不过，如果需要选择，那么康复工程师必须根据所需传递的信息类型，充分利用优于其他方式的某种感觉通道的优势。例如，在接受报警信息时，听觉往往优于视觉，因为听觉容易吸引人的注意力。表 5.7 广泛比较了听觉和视觉信息传递各自的特点。

表 5.7　听觉与视觉两种信息传递方式的比较

（摘自 Saunders 和 McCormick，1993，第 53 页表 3-1，并作修改）

适合听觉信息传递方式的情况	适合视觉信息传递方式的情况
信息简单	信息复杂
信息简短	信息冗长
事后不需查阅信息	事后可能会查阅信息
信息表示的是事件发生的时刻	信息表示的是空间位置
信息需要即刻响应	信息不需要即刻响应
病人的视觉系统负担过重	病人的听觉系统负担过重
不在设备附近的人也能接收到信息	只有非常靠近设备的人才能接收到信息
给不能阅读的人产生人工语音	如果信息包含图像那么就要用视觉方式

容错原则

康复工程学和人因学（即人机工程学）都致力于设计辅助技术，用于扩大人体的能力，并且尽可能减少出错。但是，无论设计多么完美，人为错误总是难免的。因此，在不严重影响系统的性能和安全性的前提下，辅助设备必须具备一定的容错性。各种错误大致可以分为遗漏错误、执行错误、顺序错误以及定时错误等几类。

设计良好的基于计算机的电子辅助设备必须具备下列一种或多种特性：

1）设备本身杜绝了犯错误的可能性。例如，插头只能以惟一的方式插入插座，设备能够自动拒绝不适当的操作，并给出报警等。

2）设备能够减少错误的发生，但并没有完全杜绝错误发生的可能性。例如，采

用彩色导线，并配以易于理解的接线示意图，方便用户接线。

3）设备能够减少错误造成的破坏，但不能降低错误发生的可能性。例如，采用熔丝限制过大的电流，采用机械制动器限制超范围的机械运动和速度等。

4）在菜单设计中包含“撤销”、“退出”、“返回”等命令选项。

适应性和灵活性原则

人机工程学的一个基本原则就是设备的设计应该适应使用者，而不是反过来让使用者适应设备。如果环境变化了，或者使用者掌握了技巧，能够更加熟练地操作辅助设备了，那么，设备的操作特性就必须能够随即作出相应的调整。以电子辅助交流设备为例，根据用户的需求、技能和交流环境的变化，其词汇库应该能够方便地改变。单词的选择方式和反馈方式也应该具有灵活性，例如，有时可以直接选择单词，有时又可以回到简单的行-列扫描的间接选择方式。另外，设备还应该能够给用户提供听觉、视觉，或者两者结合的反馈指示。

心理年龄和实际生理年龄相适应的原则

如果残疾人是终身残废或者严重残疾，那么，康复工程师就必须考虑病人心理年龄与实际生理年龄之间的差别。一般，先天性残疾病人由于受到各种限制，生活经历较少，常常容易模仿那些年龄比自己小得多的人的反应和行为。因此，在分析和评估辅助技术问题时，康复工程师应该首先确定服务对象的心理年龄。通过直接观察或者与家属、老师以及病人周围的社会人员交谈，可以收集残疾人的行为信息和生平信息。

为儿童和老年人设计辅助设备时，康复工程师还需要特别考虑体能因素，这些人处理和记忆各种信息的能力往往较弱。例如，从长期记忆中回忆信息、从多种响应中确定选项以及执行各种正确的响应时，儿童和老人常常需要花费较多的时间。研究结果表明，在从长期记忆中回忆信息、将注意力从一个作业转移到另一个作业以及处理概念、空间和动作之间的不一致性时，老年人的反应都要慢得多。

根据以往的经验，在为老年人设计辅助设备时，应该遵守下列指导方针：

1）增强指示信号，如增大声音、加强亮度、加大尺寸等。

2）简化各种控制和显示，以减小细节内容可能引起的混乱。

3）保持概念、空间和动作之间的高度一致性，也就是使控制、显示和设备响应之间保持良好的兼容性。

4）减少同时监视和响应多项任务的需求。

5）在两项响应任务之间留有足够的时间间隔。如有可能，就让使用者自己决定执行任务的速度。

6）提供较长的学习和练习时间。

5.5 康复工程与辅助技术的实践

5.5.1 职业发展机会

在尽可能限制医疗保健费用增长的形势下，人们有理由提出疑问，康复工程师和辅助技师是否还有就业机会。随着人口的不断老龄化，先天性智力和生理发育疾病的儿童人数不断增加，近年来国家立法的影响（见表 5.1），以及康复技术成功所带来的成本效益的体现，社会对于辅助技术的需求只会增加，不会减少。因此，致力于产品开发和辅助技术服务的技术人员的就业机会将稳步增长。

20 世纪 80 年代早期，康复工程师和辅助技师的价值还未受到广泛的重视，当时需要大量宣传教育工作。如今，虽然这方面的工作并没有完全结束，但在过去的 20 多年时间里，人们的认识已经有了很大进步。例如，联邦政府资助的加州医疗保险项目（Medi-Cal）现在就资助辅助交流设备的购买和定制。很多州政府都已经将那些能够帮助残疾人自立和获得就业机会的技术设备作为常规资助项目。

目前康复工程师和辅助技师的就业机会包括：医院的康复中心、公立学校、职业康复机构、生产厂家以及社区的康复技术服务处等，还有单独的签约服务。例如，1997 年 Virginia 州康复服务部发布的康复工程师就业广告中就有如下职业说明：

"为残疾人、职员、社区机构和雇主提供工作场所和合理接纳方面的康复工程服务和技术支持；管理、设计、改装并制造自适应设备……必须具备用于残疾人的各种技术的设计、制造和正确解决工程问题的知识和能力。具有相关设备和工具的操作技能，并能够指导他人进行辅助设备的制作。能够发展与康复工程相关的教育项目。在康复工程、机械工程或者生物医学工程等工程专业接受过正规训练，或者具备同等资历。"

另外，公告中所述的该职业的薪水和福利与其他工程职业的相当。在"康复管理"等商业杂志和 RESNA 新闻报道中经常会出现类似的就业广告。

下面以宾夕法尼亚州 Malvern 的 Bryn Mawr 康复中心为例，看一下基于医院系统的康复中心的就业机会。该中心隶属于 Jefferson 卫生系统，该系统是多家医院和长期家庭护理机构组成的非营利性组织。Bryn Mawr 康复中心提供康复工程和辅助技术服务，它有一个老年人康复诊所，专门组织几个相关部门帮助老年人，使他们能够在自己家中生活更长时间。联邦政府医疗保险 Medicare 支付该诊所为老年人独立生活进行的评估和技术服务费用。该项目之所以能够获得资助，是因为保持老年人的身体健康并让他们呆在自己家中可以节省大笔医疗费用。

康复工程师和辅助技师还可以到必须遵守美国残疾人教育法（Individuals with Disabilities Education Act）的学校里工作。这里，他们的工作是评估、改装或者定制辅助设备，帮助教学人员改建课堂设施，向资助机构提出所需的有关教育技术需求。在这种工作环境中，工程师们必须能够与教师、家长、学生和学校管理人员等非技

术人员建立良好的共事关系。

还有一个康复工程师和辅助技师的就业岗位是在社区服务机构，如当地的脑瘫协会或者国家肢障儿童协会（National Easter Seals Society）的当地分会等。通过收取服务费用、接受捐款和保险费等各项经费，社区服务中心的康复工程服务在财政上是可行的。这些中心雇用辅助技术专家完成信息咨询、评估、定制产品改进和培训等工作。

康复工程师也可以与客户单独签订合同，或者受聘于辅助设备生产厂家。由于康复工程师懂得技术原理并了解各种残疾状况的本质，因此，他们可以成为厂商与消费者之间的联系人，帮助双方确定和评估新产品。另外，成为独立咨询师的康复工程师也能够给消费者、资助机构和涉及工伤赔偿的保险公司提供相关知识和可信赖的建议。这种咨询服务通常包括提供相关辅助技术的信息、进行客户的评估、评定辅助设备的适用性等。康复工程师要成为独立的咨询师必须取得美国专业工程师（Professional Engineer，PE）执照，而且还要通过下一节将介绍的 RESNA 资格认证。获取专业工程师执照的第一步通常是要通过各州执照管理委员会设立的工程基础考试（Fundamentals of Engineering Examination）。

5.5.2 康复工程展望

康复工程作为一门独立的学科现在正处于不断成熟的发展时期，它具有明确定义的应用范围，包括应用新材料制造轮椅和假肢，为提高残疾人的独立生活能力和就业能力开发各种辅助技术，利用无线通信技术和信息技术排除各种障碍，利用数字化技术和先进的信号处理技术开发助听器和交流设备，开发神经修复产品辅助手部功能，以及呼吸、站立、甚至行走。

自从 1973 年康复法及其后续的 1992 年和 1998 年修改案颁布以来，康复工程在美国已经被看做值得许多政府机构支持的领域，很多大学都在该领域开设了正式的研究生教育计划。各种公共和私营保险公司都可以为康复工程服务项目支付费用。报纸和就业公告中经常出现康复工程师的招聘广告。1990 年，美国残疾人法案赋予残疾人公民权，并且强制要求所有雇员超过 25 名的公司都合理接纳残疾人。目前学术刊物发表的研究论文已经涉及康复工程的所有研究方向。对该领域感兴趣的学生也与日俱增。那么，接下来会怎样呢?

根据近年的发展状况，未来 10 年康复工程实践及其研究和开发工作发展的主流趋势可能有：

1）美国将完善康复工程师的资格认证体系。资格认证是非政府机构或专业协会确认某人在特定领域所具备的知识和技能的过程。RESNA 领导辅助技术提供人员的资格认证工作。RESNA 的专业康复工程师的认证条件是：申请人必须是法律上所谓的注册专业工程师（PE），具备康复技术所需的相关工作经验，并通过一个包含 200 道选择题的考试。对于非工程师人员，可以申请辅助技术技师（Assistive Technology Practitioner，ATP）或者辅助技术供货人（Assistant Technology Supplier，ATS）的

资格认证。本章末尾有 RESNA 资格认证考试的考题示例。

2）康复技术工程师和技师的教育和培训将遍及世界各地。该领域的非正式国际交流已经产生。RESNA 等政府机构和专业协会的倡议将促进这方面的发展。例如，美国教育部就资助了北美和欧洲几所大学的一项培养康复工程师计划，该项目的一个间接目标就是希望实现信息沟通、建立学生和研究人员之间的正式交流。

3）各种残疾人消费品的设计和使用将越来越普遍。消费品技术的发展已经给残疾人带来了巨大的利益。例如，语音识别系统使行动不便的病人可以利用计算机与家人和外界进行交流。远程通信使残疾人不必到特定的工作场所也能完成工作。有趣的是，原来为残疾人设计的产品现在开始反过来给大众带来了好处。例如，曾经为残疾人设计的大旋钮、橱柜把手、便利门、路缘坡、闭路字幕电视、大显示屏等，现在已成为人人欢迎的产品。未来，随着老龄人口和残疾人口的增加，这种有利于大众使用的产品将越来越普及。产品的可交换性、部件模块化、用户界面友好等通用的产品设计原则将不断推广。

4）人机工程学在康复工程中的作用将越来越显著。为残疾人设计产品时，人机关系和人体因素至关重要，常常决定了产品的成败。基于此认识，1994 年 9 月《IEEE 康复工程汇刊》出版了“康复人机工程学与人因学”专刊。人因学与人机工程学协会还专门设立了的“医疗系统与康复工程”分会。在今后 10 年里，越来越多的康复工程培训项目将提供人机工程学与人因学课程。康复工程师将普遍认识和重视人体因素，他们为残疾人设计的各种特殊产品都将结合人体因素。

5）有关康复工程服务的成本与效益分析将势在必行。在医疗领域，控制成本和提高效率已经成为人人关心的问题。联邦政府不久就会要求进行有关康复工程师和辅助技师工作的经济模式与社会效益的分析。卫生保健机构和医疗管理组织将不再只根据零散的报告来审定对于康复工程与辅助技术的资助（Gelderbom 和 de Witte，2002；Andrich，2002）。基金资助机构下一步很可能将主要倡导公正的研究人员在康复工程领域的定量化的纵向深入研究。

6）有关部门将建立各类辅助设备的质量保证标准和性能标准。随着康复工程和辅助设备服务费用的增加，毫无疑问需要客观的质量保证和技术水平评价标准。例如，RESNA 和美国国家标准学会（American National Standards Institute）正在联合建立轮椅标准委员会，辅助技术服务提供人员和供应商的资格认证工作也正在展开。

7）无线技术的应用将极大地增强残疾人的独立生活能力。例如，全球定位系统、互联网地图、手机三角基站定位、无处不在的无线电频率识别标签等导航辅助设备，将使盲人在室内、室外都可以像正常人一样方便地行走。无线技术还可以在日常生活中帮助智障病人。例如，提醒器、跟踪器、游走器以及个人资料携带器等都可以帮助病人记住约会和服药时间、确定自己所处的位置、按照常用指令行事或者寻求援助等。

8）科学技术将成为功能强大的补偿器，它可以补偿残疾人在操作、定位、行走和交流方面遇到的限制。在未来 20 年里，康复工程师将利用科技手段使残疾人能够

坐在核能动力轮椅上，通过轮椅上安装的基于互联网的语音控制无线计算机工作站，实现数据与信息操作，遥控系统的工作模式。残疾人将不必每天坐车去上班，他们可以在家里适合自己的特殊环境中工作，并可以利用辅助技术拓展自身的能力。例如，强大的语音识别系统可以将病人发音困难的模糊不清的语言实时地自动识别并转化为清晰的语言。在如今科学技术突飞猛进的形势下，这些未来的辅助设备并不仅仅是梦想，而只是可以预计的现实的延伸。

有志于康复工程与辅助技术研究和开发的学生们，在完成正规训练之后，将为加速实现上述梦想作出贡献。不过，未来康复工程师的总体职责并不会改变，他们仍然要评估病人的需求和限制条件，应用本章所介绍的各种原则来设计、改装和制造辅助设备。

习　题

就像本章所述的各种工程设计过程一样，回答下列问题时，读者可能需要查找本教材以外的资料。建议参考本章末尾的推荐阅读资料，也可以利用互联网查找所需信息。

1. 康复工程和辅助技术领域深受美国联邦政府立法的影响，请阐述政府立法对于康复工程技术人员以及辅助技术市场的发展所产生的影响，并举例说明。

2. 假设你是在学校工作的康复工程师，有位教师要求你为一名二年级的8岁女生设计并制作一种器具，使她能够对课堂提问作出响应或者提出自己的请求。该女生坐电动轮椅，患有多重残疾，上肢不能做什么动作，说话也表达不清。请列出一份定性和定量问题分析清单，并给这种信号发生装置设定一套性能指标。

3. 假如需要设计一种语音输出示波器，用于视力受损电子工程专业学生的运算放大器实验，请问这种示波器需要具备哪些特性？请列出其性能指标。

4. 假设需要设计一种安装在手动轮椅上的可折叠膝上台板，请写出其性能指标。提示，可以从以下几方面考虑：台板的最大和最小尺寸范围应该为多少？它需要承受多少重量？加上台板后会增强还是减弱轮椅使用者的独立性？台板应该采用什么材料制作？

5. 请给例5.1所述的安装在电动轮椅上的下肢升降器画一张草图，画出这种轮椅的侧视图和后视图。

6. 请查找符合下列性能指标的电子交流产品：语音输出，印有图标的薄膜式键盘，便携式，重量不超过7.5lb，厚度小于2.5in，尺寸不大于标准三孔活页夹，并且能够根据用户的要求快速更改常用的词语。提示：可以查询“The Closing the Gap Product Directory”和“Cooperative Electronic Library on Disability”等产品目录，后者在美国Madison的Wisconsin大学的Trace研究发展中心（Trace Research and Development Center）网站上可以找到。

7. 残疾人的残疾情况和能力通常取决于其医学上的生理状况。

a）患有 C5-C6 部位脊髓损伤（Spinal Cord Injury，SCI）的病人具有怎样的运动和感觉能力和限制？

b）患有多发性硬化症的病人的运动和感觉能力和限制又如何？并请比较这两类病人的预后情况。

8. 便携式电子辅助设备必须用电池供电。请根据你所学过的有关技术手册和电池产品手册，比较一次性碱性电池和可充电的锂电池的优点和缺点。请分析它们的电池容量、重量和充电电路等技术问题，还有成本问题，以及使用者的爱好、电池装错和充电器误用的潜在风险、用户的便利等使用问题。

9. 假设有位截瘫的年轻人想重新滑雪、驾驶帆船、打高尔夫球。请分别给每种运动列出 4～5 种可能需要的运动器具改装方案，并画出草图。

10. 有位患有肌营养不良症的 21 岁女性病人希望通过辅助设备使用计算机、进行写作、使用电子数据表、玩计算机游戏等。除了手腕和手指有一些运动能力以外，她的四肢几乎都不能动。她的左手大约能够伸展 6in 距离，右手能够伸展 2in 距离，两只手一共能够伸展 8in 范围。如果将手腕托住，那么，她两只手的食指就都能够很好地控制。请根据这些情况画出她双手能够达到的工作范围，并说明她的标准计算机键盘或者缩小的键盘需要进行哪些改装。其他有关资料请查询“The Closing the Gap Product Directory”和“Cooperative Electronic Library on Disability”，以及本章末尾所列的推荐阅读资料。

11. 计算机的用户界面主要有两种，一是 UNIX 等操作系统使用的命令行界面（Command Line Interface，CLI），二是 Windows XP 或者苹果 OS X 等操作系统使用的图形用户界面（Graphical User Interface，GUI）。对于行动不便的残疾人而言，两种界面各有优缺点。请比较 CLI 和 GUI 的优点和缺点，并说明在什么情况下、对何种残疾人 CLI 优于 GUI?

12. 辅助交流设备是一类主要的辅助设备。请描述文本转换成语音的过程中需要哪些电子数据处理步骤，如何利用个人计算机的先进技术提高转换速度，并使语音输出更加逼真?

13. 在例 5.3 中，如果此人将双腿同时垂直竖起，且 D 为 14in，那么，秤的第二次读数 S_2 将是多少?

14. 在例 5.4 中，如果砝码以 1.5ft/s^2 的加速度下落，那么，施加在滑轮上的拉力是多少?

15. 如图 5.10 所示，某人手上托着 25lb 重物，重物的中心与肘关节之间的距离为 14in。假设屈肌的作用力方向与前臂呈 90°角，且力的作用点距离肘关节 2in，前臂重量为 4.4lb；请问此时屈肌需要产生多大的收缩力。提示：利用静力学平衡方程

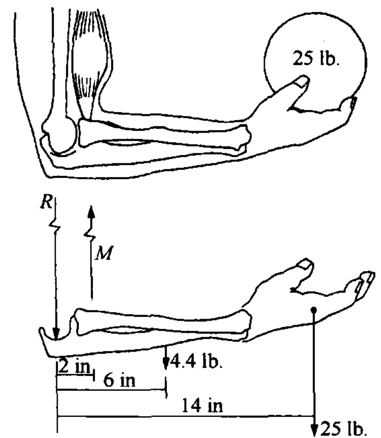


图 5.10　肘部屈肌锻炼时肘关节的静态作用力（摘自 Le Veau，1976）

$\sum F_x = \sum F_y = \sum M = 0$。

16. 如图 5.11 所示，当体重为 200lb 的人单腿站立时，他的股骨头承受的作用力是多少？提示：利用静力学平衡方程$\sum F_x = \sum F_y = \sum M = 0$，并结合图 5.11 所示的骨骼受力图进行分析。

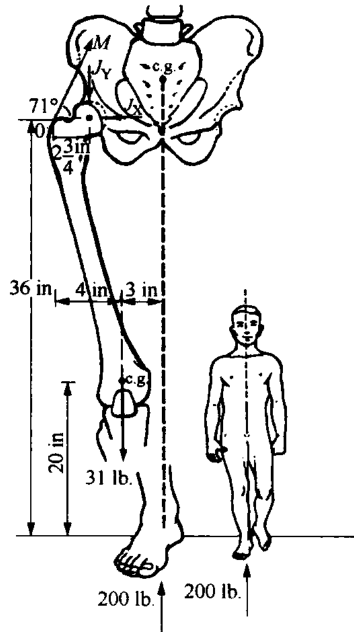


图 5.11 单腿站立时股骨头承受的压力情况（摘自 Le Veau，1976）

17. 在静止或者匀速运动状态下，如果轮椅和乘坐者整体的重心（Center of Gravity，CG）在垂直方向上的投影落在后轮与斜坡接触点之后，那么，轮椅就会向后翻倒。如图 5.12 所示，向后翻倒的临界角 θ_r 取决于 CG 与轮椅后轮车轴之间的水平距离 d_1 和垂直距离 d_2。

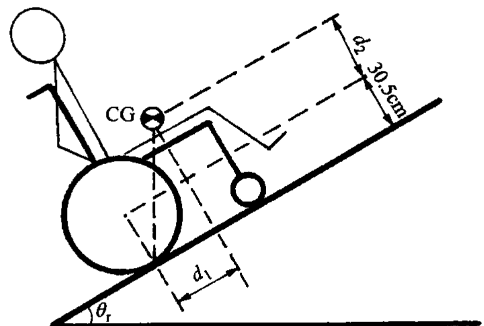


图 5.12 轮椅将向后翻倒时的临界状态

注：当 CG 在垂直方向上的投影落在后轮与斜坡接触点之后时，轮椅就会向后翻倒（摘自 Szeto 和 White，1983）。

a）请利用静力学方法推导 θ_r、d_1 和 d_2 之间的关系式。

b）应用图 5.7 所示的平板法设计一种确定 d_1 的方法。

c）假设对于形体完整的人 d_1 和 d_2 的平均值分别为 13cm 和 24cm，那么，θ_r 应为多少？

d）如果轮椅乘坐者向前倾，而不是背靠轮椅，那么，d_1 和 d_2 将如何变化？θ_r 又

会如何变化?

18. 请分析图 5.13 所示的静力学状态，并使用图中的数据推导轮椅向前翻倒的临界角 θ_r的方程。假设 d_1和 d_2的值与习题 17 相同，d_4和 d_5的平均值分别为 27cm 和 49cm，那么，θ_r的值为多少？其中，d_1和 d_2分别为 CG 与后轮车轴之间的水平和垂直距离，d_4和 d_5分别为 CG 与前脚轮轴心之间的水平和垂直距离。

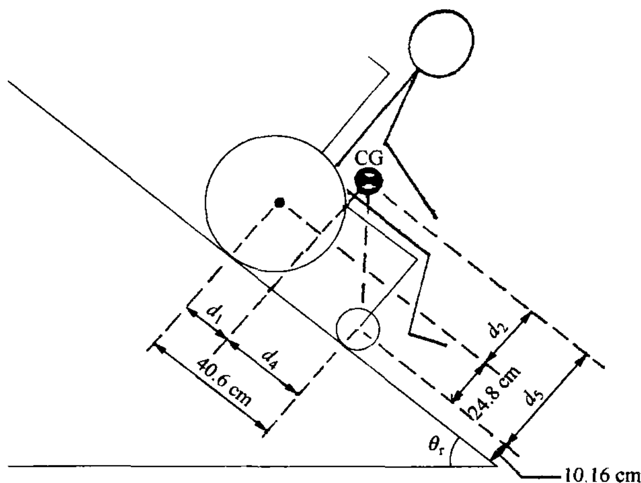


图 5.13 轮椅将向前翻倒时的临界状态

注：当 CG 在垂直方向上的投影落在前脚轮与斜面接触点之前时，轮椅就会向前翻倒（摘自 Szeto 和 White，1983）。

19. 只有头部可以控制的残疾人可以用 Prentke Romich 公司（Prentke Romich Co.，Wooster，OH）的产品 HeadMaster 来操作计算机。HeadMaster 模仿鼠标输入信号的工作原理，它由一个通过导线与计算机相连的头套构成。头套上有一个传感器，可以检测头部运动，并将此运动转化为类似于鼠标移动的二维信号。头套上还有一个吹吸式的气动开关，相当于鼠标的按键。请根据这些有关 HeadMaster 的介绍，画出该设备的工作原理框图，以及所需的基本器件，包括超声信号源、检测器、定时器、信号处理器等。

20. 请根据表 5.8～表 5.10 所列字母使用频率的数据，设计一个用于交流设备行-列扫描的优化字母阵列。行-列扫描时，先从上至下逐行点亮字母阵列的每一行，当闪亮的行包含所需字母时，用户按下开关选择该行；然后再在所选行中从左至右进行扫描，当所需字母闪亮时，用户再按一次开关。这样，第一次按下开关选择的是行，第二次按下开关选择的是列。提示：将最常用的字母排在最先扫描到的位置。请参考例题 5.5。

表 5.8 英文课本 10000 个字母中单字母出现频率的统计数据（摘自 Webster 等人，1985）

E=1231	L=403	B=162
T=959	D=365	G=161
A=805	C=320	V=93
O=794	U=310	K=52
N=719	P=229	Q=20
I=718	F=228	X=20
S=659	M=225	J=10
R=603	W=203	Z=9
H=514	Y=188	

表 5.9 英文课本 25000 个字母中常用双字母和三字母组合的出现频率统计数据

（摘自 Webster 等人，1985）

双字母组合				
TH=1582	HE=542	ON=420	NT=337	RA=275
IN=784	EN=511	OU=361	HI=330	RO=275
ER=667	TI=510	IT=356	VE=321	LI=273
RE=625	TE=492	ES=343	CO=296	IO=270
AN=542	AT=440	OR=339	DE=275	
三字母组合				
THE=1182	ERE=173	HAT=138	NCE=113	MAN=01
ING=356	HER=170	ERS=135	ALL=111	RED=101
AND=284	ATE=165	HIS=130	EVE=111	THI=100
ION=252	VER=159	RES=125	ITH=111	IVE=96
ENT=246	TER=157	ILL=118	TED=110	
FOR=246	THA=155	ARE=117	AIN=108	
TIO=188	ATI=148	CON=114	EST=106	

表 5.10 英文课本 242432 个单词中常用单词出现频率的统计数据

（摘自 Webster 等人，1985）

THE=15568	FOR=1869	HAVE=1344	THIS=1021
OF=9757	AS=1853	YOU=1336	MY=963
AND=7638	WITH=1849	WHICH=1291	THEY=959
TO=5739	WAS=1761	ARE=1222	ALL=881
A=5074	HIS=1732	ON=1155	THEIR=824
IN=4312	HE=1721	OR=1101	AN=789
THAT=3017	BE=1535	HER=1093	SHE=775
IS=2509	NOT=1496	HAD=1062	HAS=753
I=2292	BY=1392	AT=1053	WHERE=753
IT=2255	BUT=1379	FROM=1039	ME=752

21. 电子导盲犬是盲人行走的电子辅助设备。请列出这种设备应该具备的特殊功

能和信息处理步骤，尽可能给出细节并予以说明。提示：可以考虑障碍物检测、信息指示、自我驱动系统、惯性制导、路线记忆、电源等问题。

22. 用户通过视觉活动从给定的一组选项中作出正确的选择，这是很多辅助设备的基础。其中的视觉活动有视觉追踪和视觉扫描两种，请根据动眼机制分析这两种视觉活动之间的差别。

23. 根据表 5.7 的分析，下列情况分别适于使用哪种语音合成技术：能够给会拼写的病人提供无限词汇量的辅助交流系统；能够给盲人朗读整个计算机屏幕上所显示内容的语音输出系统；给只需要有限词汇量的女孩设计的辅助交流系统；环境控制设备的语音反馈系统，可以返回单个单词命令，如"开"、"关"、"灯"、"床"、"电视"和"窗帘"等。请解释并论证你的答案。

24. 严重视觉受损病人安全且独立的行走至今仍然是一个难题。为了减少这些病人对导盲犬和明眼人的依赖性，人们开发了各种使用全球定位系统（Global Positioning System，GPS）接收器的便携式导航辅助系统，并且已经面市。

a）请在互联网上搜索基于 GPS 的盲人便携式导航辅助产品，并回答以下问题：GPS 的工作原理是什么？在任何地方都能够可靠地接收 GPS 信号吗？GPS 信号的分辨率有多高？这种分辨率水平是否足以找到建筑物的入口？当 GPS 信号丢失时，航迹推算和惯性制导是否能起作用？

b）请阐述理想便携式盲人导航系统所需的各种操作功能。可以考虑用户界面、输入输出需要、产品零售价格等问题，并列举所需考虑的人机工程学设计问题。

RESNA 辅助技术资格认证考试选择题样题

1. 下列哪些能力是建立娴熟的上肢运动所必需的？

a. 站姿的平衡反应；

b. 跨越中线的能力；

c. 良好的躯干和头部姿势控制能力；

d. 手指抓握能力。

2. 某位患有 Duchenne 肌营养不良症的 12 岁男孩准备接受行走辅助系统，临床医生注意到他的躯干向左侧弯曲，那么，下一步首先应该确定的是该病人是否患有：

a. 脊柱后凸；

b. 脊柱前凸；

c. 左侧虚弱；

d. 脊柱侧凸。

3. 培训和指导辅助技术设备使用的最合适的场所是：

a. 没有干扰的安静地方；

b. 使用者的家中；

c. 设备将使用的环境；

d. 可以找到多位临床医生的培训中心。

4. 患有 C4-C5 四肢瘫痪的建筑师想利用计算机辅助设计系统（CAD）重新开展工作，那么，首先要确定该用户可以使用下列哪种工具?

a. 口控操纵杆；

b. 眨眼控制开关；

c. 鼠标输入替代物；

d. 吹吸开关。

5. 根据残疾人教育法，辅助技术设备的定义是：

a. 能够增强人体功能的设备；

b. 能够改善行走或者交流的设备；

c. 能够补偿身体或感觉功能损伤的设备；

d. 耐用的医疗设备。

6. 临床医生的医疗报告中除了诊断信息以外还必须包括：

a. 辅助技术所需费用；

b. 病人的预后；

c. 病人的活动能力；

d. 病人的肌张力。

7. 塑料是大小便失禁病人坐具的理想材料，这是因为这种材料：

a. 重量轻；

b. 比木材便宜；

c. 不吸水；

d. 可以与轮椅分离。

8. 如果想改造新买的辅助设备产品的结构，那么最重要的是必须考虑：

a. 将来其他人的使用；

b. 放弃保修；

c. 转售的价值；

d. 产品的外观。

9. 某客户想通过语音识别系统使用计算机，下列哪种因素对于这种方法的成功实现最不重要?

a. 手的功能；

b. 声音的清晰度；

c. 语音识别系统使用的培训；

d. 所用计算机系统的类型。

10. 有一位 9 岁女孩，两年前接受培训，并开始使用新购买的电动轮椅，但现在她不能再驾驶轮椅了。在调查其中的原因时，首先必须：

a. 与孩子和她父母进行面谈；

b. 对孩子进行认知能力的评估；

c. 重新测试轮椅的性能；

d. 联系轮椅生产厂家。

推荐阅读资料

Adams, R.C., Daniel, A.N., McCubbin, J.A. and Rullman, L. (1982). *Games, Sports, and Exercises for the Physically Handicapped,* 3rd Ed. Lea & Febiger, Philadelphia.

American Academy of Orthopaedic Surgeons. (1975). *Atlas of Orthotics: Biomechanical Principles and Application.* Mosby, St. Louis.

Andrich, R. (2002). The SCAI instrument: Measuring costs of individual assistive technology. *Technol. & Disabil.* **14(3)**: 95–99.

Bailey, R.W. (1989). *Human Performance Engineering,* 2nd Ed. Prentice Hall, Englewood Cliffs, NJ.

Blackstone, S. (Ed.). (1986). *Augmentative Communication: An Introduction.* American Speech-Language-Hearing Association, Rockville, MD.

Childress, D.S. (1993). Medical/technical collaboration in prosthetics research and development. *J. Rehab. R&D* **30(2)**: vii–viii.

Church, G. and Glennen, S. (1992). *The Handbook of Assistive Technology.* Singular, San Diego, CA.

Cook, A.M. and Hussey, S.M. (2002). *Assistive Technology: Principles and Practice,* 2nd Ed. Mosby, St. Louis.

Cooper, R.A., Robertson, R.N., VanSickel, D.P., Stewart, K.J. and Albright, S.J. (1994). Wheelchair impact response to ISO test pendulum and ISO standard curb. *IEEE Trans. Rehab. Eng.* **2(4)**: 240–246.

Daus, C. (1996, April/May). Credentialing rehabilitation engineers. *REHAB Management,* 115–116.

Deatherage, B. (1972). Auditory and other sensory forms of information presentation. In H. Van Cott and R. Kincade (Eds.), *Human Engineering Guide to Equipment Design.* Government Printing Office, Washington, DC.

Department of Rehabilitative Services, Commonwealth of Virginia. (1997, Spring). Job Announcement for a Rehabilitation Engineer.

Galvin, J.C. and Scherer, M.J. (1996). *Evaluating, Selecting, and Using Appropriate Assistive Technology.* Aspen, Gaithersburg, MD.

Gaines, H.F. (1940). *Elementary Cryptanalysis.* Chapman and Hall, London.

Gelderblom, G.J. and de Witte, L.P. (2002). The assessment of assistive technology outcomes, effects and costs. *Technol. & Disabil.* **14(3)**: 91–94.

Goldenson, R.M., Dunham, J.R. and Dunham, C.S. (1978). *Disability and Rehabilitation Handbook.* McGraw Hill, New York.

Guthrie, M. (1997, July). Where will the jobs be? *TeamRehab Rep.,* 14–23.

Hammer, G. (1993, September). The next wave of rehab technology. *TeamRehab Rep.,* 41–44.

Jones, M., Sanford, J. and Bell, R.B. (1997, October). Disability demographics: How are they changing? *TeamRehab Rep.,* 36–44.

Lange, M. (1997, September). What's new and different in environmental control systems. *TeamRehab Rep.,* 19–23.

Le Veau, B. (1976). *Biomechanics of Human Motion,* 2nd Ed. Saunders, Philadelphia.

McLaurin, C.A. (1991, Aug/Sep). A history of rehabilitation engineering. *REHAB Manag.,* 70–77.

Mercier, J., Mollica, B.M. and Peischl, D. (1997, August). Plain talk: A guide to sorting out AAC cevices. *TeamRehab Rep.,* 19–23.

National Institute on Disability and Rehabilitation Research (NIDRR). (1999, December).

Long Range Plan 1999–2003, Washington, DC.

Philips, B. and Zhao, H. (1993). Predictors of assistive technology abandonment. *Assistive Technol.* 5(1): 36–45.

Reswick, J. (1982). What is a rehabilitation engineer? *Ann. Rev. Rehabilitation* 2.

Sanders, M.S. and McCormick, E.J. (1993). *Human Factors in Engineering and Design,* 7th Ed. McGraw-Hill, New York.

Scherer, M. (1993). *Living in the State of Stuck: How Technology Impacts the Lives of People with Disabilities.* Brookline Books, Cambridge, MA.

Smith, R.V. and Leslie, J.H., Jr. (Eds.). (1990). *Rehabilitation Engineering.* CRC, Boca Raton, FL.

Stolov, W.C. and Clowers, M.R. (Eds.). (1981). *Handbook of Severe Disability.* U.S. Department of Education, Washington, DC.

Szeto, A.Y.J., Allen, E.J. and Rumelhart, M.A. (1987). "Employability Enhancement through Technical Communication Devices," *American Rehabilitation,* Vol. 13(2):8–11 & 26–29, April-June.

Szeto, A.Y.J. and Riso, R.R. (1990). "Sensory Feedback using Electrical Stimulation of the Tactile Sense," in *Rehabilitation Engineering,* R.V. Smith and J.H. Leslie, Jr. (Eds.), CRC Press, pp. 29–78.

Szeto, A.Y.J., Valerio, N. and Novak, R. (1991). "Audible Pedestrian Traffic Signals, Part 1. Prevalence and Impact, Part 2. Analysis of Sounds Emitted, Part 3. Detectability," *Journal of Rehabilitation R & D,* 28(2):57–78.

Szeto, A.Y.J., Allen, E.J. and Littrell, M.C. (1993) "Comparison of Speed and Accuracy for Selected Electronic Communication Devices and Input Methods," *Augmentative and Alternative Communication,* Vol. 9, Vol. 4, December, pp. 229–242.

U.S. Census Bureau. (1990). *National Health Interview Survey on Assistive Devices (NHIS-AD).* Washington, DC.

Webster, J.G., Cook, A.M., Tompkins, W.J. and Vanderheiden, G.C. (Eds.). (1985). *Electronic Devices for Rehabilitation.* Wiley Medical, New York.

Wessels, R., Dijcks, B., Soede, M., Gelderblom, G.J. and De Witte, L. (2003). Non-use of provided assistive technology devices: A literature overview. *Technol. & Disabil.,* Vol. 15, 2003, 231–238.

Wilcox, A.D. (1990). *Engineering Design for Electrical Engineers.* Prentice Hall, Englewood Cliffs, NJ.

第 6 章　生物材料学

Liisa T. Kuhn 博士[⊖]

本章目录

⊖ Katharine Merritt 和 Stanley A. Brown 也参与了本章的撰写工作。——原文注

本章主要内容：

- 天然组织结构的复杂性；
- 机体对于植入材料的几种不同生物反应；
- 各种医用材料的特性和优势；
- 医疗器械仿生特性的设计；
- 可以增强或者改变生物材料与细胞之间相互作用的材料修饰方法；
- 各种组织工程支架的制备方法；
- 如何寻找证明医疗产品安全性的现成测试方法。

6.1 医用材料——从修复到再生

多年来，医用材料（又称生物材料）在人体创伤和疾病的治疗过程中发挥着巨大的作用。19世纪后期以来，尤其是19世纪60年代Joseph Lister医生发明了无菌手术之后，生物材料的应用得到了快速发展。早在19世纪末至20世纪初就有人用金属器械修复骨折；1938年首例全髋关节置换手术取得成功；到了20世纪50～60年代，又出现了聚合物人工角膜和血管置换术。如图6.1所示，如今，生物材料的使用几乎已经遍及人体全身每个部位。据估计，2002年美国使用的生物材料类医疗器械的数量是：

- 全髋关节置换体448000个；
- 膝关节置换体452000个；
- 肩关节置换体24000个；
- 牙科移植体854000个；
- 冠状动脉支架1204000个；
- 冠状动脉导管1328000个。

每年，生物材料的应用都拯救了数百万人的生命，同时还改善了另外数百万人的生活质量。但是，这个领域仍然存在着广阔的研究空间和创新空间；因为没有一

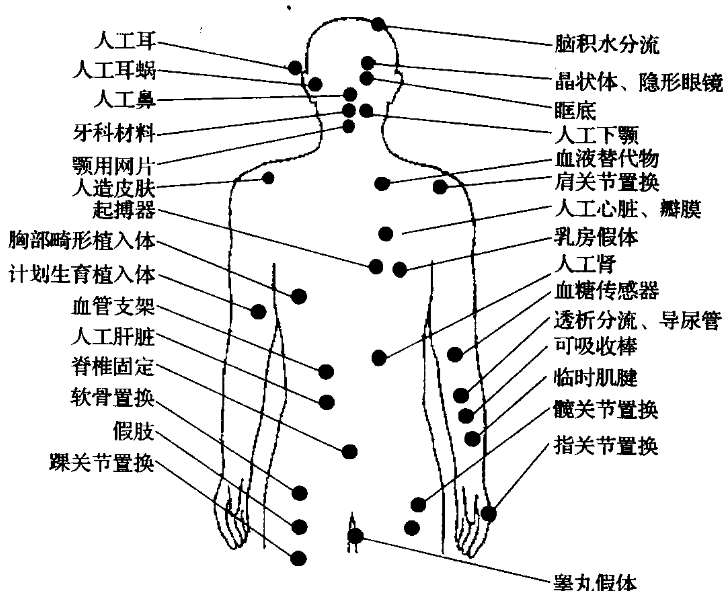


图 6.1 生物材料在创伤和疾病治疗中发挥着巨大的作用，其应用已遍及人体全身

种材料可以胜任所有生物材料的各种不同的应用要求，并且随着医学的发展，新的应用需求仍然在不断涌现。而且，生物机体对生物材料产生的很多反应还没有得到解决，生物材料在组织再生中能够发挥的作用正在研究之中，这些问题都不断激励着生物材料的研究和新产品的开发。

在过去很长的一段历史时期，人们对生物组织的机制缺乏深入的认识，以为生物医学工程技术只要利用简单的生物材料就能够完全替代已经丧失功能的生物组织。随着对生物组织、疾病和创伤的认识不断加深，人们才开始逐渐形成要修复损坏组织的观念。近些年来，随着干细胞研究的发展，医学界相信，通过基于细胞的组织工程技术可以再生病变和损坏的生物组织（参见本书第 7 章）。于是，随着不断改变的医学理念，有关生物材料的观念已经发生了演变。1987 年 Williams 把生物材料定义为“用于医疗器械并与生物系统具有相互作用的无生命材料”。这个定义如今仍然正确，它既可以包括早期的生物材料替代物，如木头腿、玻璃眼球等；又可以包括目前再生医学中使用的生物可降解支架等材料，这些支架在组织工程中可用于输送细胞。然而，虽然定义没有改变，但是人们对于生物材料与生物系统（即人体）之间相互作用的认识却产生了深刻的变化。人们原来设想生物材料在体内具有相对惰性的性质，而现在却认为这些材料具有“生物活性”，并且可以辅助生物体再生。生物活性材料植入人体之后能够诱发生物反应，比如细胞粘附、细胞增殖等，更令人惊喜的是干细胞分化还可以导致损毁组织或整个器官的再生。

由于细胞和组织对生物材料会产生各种复杂的反应，因此，在设计、选择、合成和制造生物材料时，应该遵循自然规律办事，这种方法称为仿生学。在生物

材料学中，仿生学包括对于天然材料或者活体组织各个方面的模仿，如化学特性、微观结构、制造方法等。不过，由于天然组织的很多性能还没有被彻底了解，因此，仿生学的方法并非总能获得理想的结果。而且，对于不同的生物医学应用场合，生物材料的各种特性会有很大的差异。因此，本章除了讲述通过改变生物材料的化学和结构等特性来引导组织修复的常用方法以外，还将介绍几种主要人体器官修复以及药物输送等特殊应用的生物材料设计方法。另外，本章还将介绍医疗产品的各种标准及其管理机构，它们在确保医疗产品的安全性和有效性方面起着重要的作用。

6.2 生物材料的特性、种类以及应用

6.2.1 力学特性及其测试方法

在论述各种材料及其与生物组织之间的相互作用时，有必要先介绍一些描述材料的力学特性的基本概念。测试材料力学特性的最常用方法是拉伸试样并同时测量其作用力和变形。另外，还要进行抗压和抗弯曲试验，这两种试验所涉及的基本测量参数相同，只是数学计算公式不同。为了便于不同实验室测试结果之间的比较，有关部门制定了标准测试方法。生物材料领域中应用的大多数测试方法由美国材料与测试协会（American Society for Testing and Materials，ASTM）制定。例如，有关金属拉伸试验的条例是 ASTM E8，橡胶材料拉伸试验则是 ASTM D412，硬塑料的拉伸试验是 ASTM D638。这些方法都具体规定了试样的形状、尺寸、测试条件、计算方法和试验报告的内容等。

下面举例说明根据 ASTM E8 拉伸试验条例测试“狗骨头”形状试样的过程。这种试样两头粗、中间细。将试样两端固定在夹具上，较细的中段就是受试部位。这个中段被标记为“标距”（gage length），它表示试样形变的测量范围。力学试验机通过旋转螺纹或者液压机构来拉伸试样。力的单位是牛顿（N），试样拉伸产生的变形以毫米（mm）为单位计量。由于用于试验的试样尺寸有所不同，测量值必须归一化，以消除尺寸不同带来的影响。应力 σ 等于作用力除以试样的原始截面积，其单位是 N/m^2（即帕斯卡）；应变 ε 等于试样长度的变化量除以其原始长度，单位是%，即

$$\sigma = \text{作用力}/\text{原始截面积} \tag{6.1}$$

$$\varepsilon = [(\text{变形后的长度} - \text{原始长度})/\text{原始长度}] \times 100\% \tag{6.2}$$

根据测试结果可以作出如图 6.2 所示的应力-应变曲线，还可以计算出一组表示材料特性的数据。图中 A 区是曲线的弹性区。如果金属材料上施加的应力较小，例如图中点（1）处的应力，那么它就会产生弹性变形，一旦应力消失，它会回复到原来的长度，就像橡皮筋一样。应力-应变曲线弹性区的斜率表示材料的刚度，称为弹性模量 E，又称杨氏模量。即

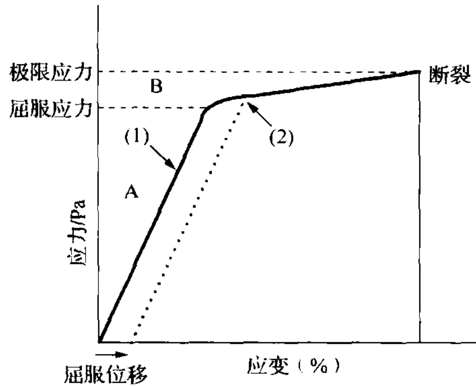


图 6.2 金属断裂之前拉伸和变形的典型应力-应变曲线

注：应力的单位是 N/m^2（即 Pa），而应变则表示为相对于原始长度的百分数。导致材料发生永久性变形的最小应力称为屈服应力（YS）。极限应力表示材料断裂前所能承受的最大应力。材料最终被破坏时的应力称为断裂强度（FS）。曲线上的 A 区是弹性区，此区域中应变的增加直接与外加应力成正比。如果外加应力很小［如图中点（1）处］，那么，一旦应力被撤除，材料就会回复到原有长度。曲线上的 B 区是塑性区，此区域中应变的变化不再与外加应力的变化成比例，此时的应力会导致材料的永久性变形。如果外加应力到达曲线上的位置（2），那么，应力被撤除之后，材料将会沿着虚线返回基线处，并永久保留图中箭头所示的形变。

$$E=\sigma/\varepsilon \text{ 的初始斜率}=\text{应力/应变} \tag{6.3}$$

随着外加应力的增加，曲线不断上升，到达某个点之后，金属就会开始产生永久性变形，这个点上的应力称为屈服应力（Yield Stress，YS）。如果在图中（2）处将应力撤除，那么，应力-应变记录数据就会沿着平行于弹性区的虚线返回。此时所产生的永久变形称为屈服位移（offset yield）。由于材料真正的屈服点很难测定，因此，经常用屈服位移来代替。一般，金属的屈服点定义为 0.2%应变处，而塑料的屈服点则定义为 2%应变处。产生永久变形后，如果再重新加载，那么金属的应力-应变就会沿着具有屈服位移的虚线上行，到达曲线的上段，并且在不断增加的应力作用下其应变会不断增加，这个区域称为塑性区。曲线能够到达的最大应力称为极限应力（ultimate stress）。最终金属将在断裂强度（Fracture Strength，FS）处被损坏，此时可以计算最终的应变量，即材料伸长或者缩短的百分比数值。

例 6.1 小块边长为 7mm 的立方体骨头块接受压力试验，测试时的步进压缩量约为 0.05mm，每次形变所需施加的作用力如下表所示。请画出其应力-应变曲线，并计算这块骨试样的弹性模量和极限应力。

变形/mm	压力/N	变形/mm	压力/N
0.00	0	0.36	1259.9
0.10	67.9	0.41	1190.9
0.15	267.6	0.46	1080.8
0.20	640.2	0.51	968.6
0.26	990.2	0.56	814.2
0.31	1265.1		

解：计算步骤如下：

（1）计算此立方体的截面积，即 $0.007m\times0.007m=49\times10^{-6}m^2$。

（2）根据应力 σ 等于作用力除以试样原始截面积的计算公式，计算对应于每个测量点的应力。例如，立方体被压缩 0.10mm 时的应力为 $67.9N/0.000049m^2=1.39MPa(1N/m^2=1Pa)$。

（3）计算每个测量点处的应变。例如，当立方体被压缩 0.10mm 时，其长度是 7mm－0.10mm＝6.9mm，于是，其应变就是 [(6.9mm－7mm)/7mm]×100%＝－1.43%。其实直接用压缩变形量除以原始长度就可以得到该数值。负号可以省略，它只表示试样是受压而不是受拉。计算结果列表如下：

应变（%）	应力/MPa	应变（%）	应力/MPa
0.00	0	5.14	25.71
1.43	1.39	5.86	24.30
2.14	5.46	6.57	22.06
2.86	13.07	7.29	19.77
3.71	20.21	8.00	16.61
4.43	25.82		

（4）图 6.3 是根据以上数据画出的应力-应变曲线。图中用线性回归法计算了曲

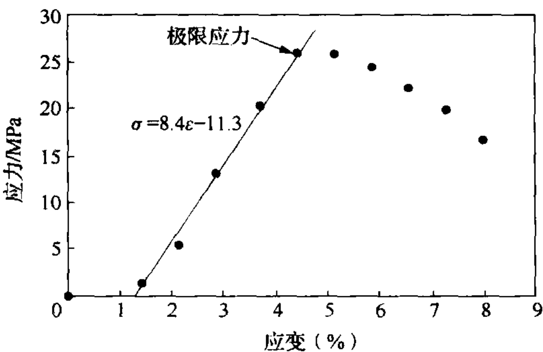


图 6.3 根据例 6.1 骨头试样的试验数据得到的应力-应变曲线

注：图中画出了利用线性回归法得到的应变 1.43%～4.43%的数据点（即线性区）的最佳拟合直线，直线的斜率为 8.4，它就是骨头的弹性模量。该材料的极限应力是 25.82MPa。

线弹性区的直线拟合。可见，其弹性模量（即直线的斜率）为 8.4，极限应力为 25.82MPa。■

描述力学试验结果的参数还有刚度、韧度、硬度等。材料弹性区的斜率（即弹性模量）也常被称为刚度。如图 6.4 所示，如果材料在断裂之前可以被拉伸很大的长度，那么它就属于延性材料；如果材料在断裂前变形不明显，那么它就是脆性材料。图 6.4 曲线下的面积具有能量单位，被称为韧度。材料的强度并不能直接从应力-应变曲线上求得，但是它与材料的硬度相关。强度越高的材料一般就越硬。材料的硬度可以通过作用力已知的锐器落在材料表面上所造成的凹痕来测量。硬度可能是衡量材料耐磨性的最重要的特性。

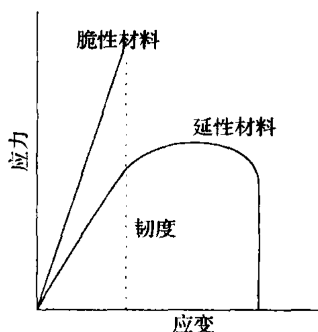


图 6.4　脆性材料在很小的应变下就会断裂，而延性材料可以在断裂前产生很大的伸展或压缩。应力-应变曲线下的面积表示韧度

材料还有一个特性就是疲劳强度（fatigue strength），也称为疲劳极限（endurance limit），它在图 6.4 中没有表示出来。如果像图 6.2 所示那样反复测试材料，每次施加的应力到达点（2）处就撤除，材料产生永久变形。这种重复试验的作用就像来回弯折回形针一样，最终材料总会断裂。但是，如果施加的应力只到达点（1）就撤除，那么，材料不会产生永久变形；反复加载到（1），也不会断裂。对于某些金属，在某个应力水平下，理论上可以反复加载和卸载无穷多次，仍然不会受到破坏。但是，实际上，疲劳极限试验只进行一定次数的交变负荷测试，如 10^6 次或 10^7 次。显然，疲劳极限是承载性医疗器械设计中的一个关键指标，如全髋关节置换部件每年平均要经受百万次负荷变化，而人工心脏瓣膜每年平均要承受 4000 万次负荷变化。

6.2.2　金属

金属作为生物材料具有高强度、高韧性和抗腐蚀等优点。表 6.1 和表 6.2 列出了医疗器械中的一些常用金属及其力学特性。很多矫形器都用金属制作，如髋关节和膝关节置换体等（见图 6.5 和图 6.6）。这些植入体可以减轻疼痛并恢复原始软骨被磨损的关节的功能。如图 6.7 所示，骨折治疗期间用于固定骨头的接骨板和螺钉也都用金属制作。有时康复后金属板和螺钉会被取出，但某些情况下它们会一直留在

体内。金属器械还可以用于替代损坏的脊椎骨，将脊椎连接起来（见图 6.8），也可以用作牙根植入体（见图 6.9）。

表 6.1 各种材料及其医学应用

材料类型	用途
金属	
不锈钢	关节置换体、骨折固定、心脏瓣膜、电极
钛和钛合金	关节置换体、牙桥、牙科植入体、冠状动脉支架
钴铬合金	关节置换体、骨折固定
金	牙科填料、牙冠、电极
银	起搏器导线、缝合线、牙科用汞齐合金
铂	电极、神经刺激器
陶瓷	
氧化铝	髋关节植入体、牙科植入体、耳蜗置换体
氧化锆	髋关节植入体
磷酸钙	骨移植片替代物、全关节置换体表面涂层、细胞支架
硫酸钙	骨移植片替代物
碳	心脏瓣膜涂层、矫形植入体
玻璃	骨移植片替代物、牙科填料
聚合物	
尼龙	手术缝合线、人工肠胃、气管导管
硅胶	指关节、人造皮肤、乳房植入体、人工晶状体、导尿管
聚酯	可吸收缝合线、骨折固定、细胞支架、皮肤创口覆盖物、药物输送载体
聚乙烯（PE）	髋关节和膝关节植入体、人工肌腱和人工韧带、人造血管、假牙、面部植入体
聚甲基丙烯酸甲酯（PMMA）	骨水泥、人工晶状体
聚氯乙烯（PVC）	导管、面部假体
天然材料	
胶原和凝胶	美容外科、创面敷料、组织工程、细胞支架
纤维素	药物输送
甲壳素	创面敷料、细胞支架、药物输送
陶瓷或脱矿陶瓷	骨移植片替代物
藻酸盐	药物输送、细胞封装
透明质酸	术后粘连的防治、药物输送、细胞支架

表 6.2　有关文献报道的某些材料的力学特性

	MPa	MPa	%	GPa
金属				
高强度碳钢	1600	2000	7	206
F138[①]，退火	170	480	40	200
F138，冷处理	690	860	12	200
F138，金属丝	—	1035	15	200
F75[②]，铸造	450	655	8	200
F799[③]，锻制	827	1172	12	200
F136[④]Ti64	795	860	10	105
金		2～300	30	97
铝，2024－T4	303	414	35	73
聚合物				
PEEK		93	50	3.6
PMMA 铸造		45～75	1.3	2～3
乙缩醛（POM）		65	40	3.1
UHMWPE		30	200	0.5
硅橡胶		7	800	0.03
陶瓷				
氧化铝		400	0.1	380
部分稳定氧化锆、氧化镁		634		200
全稳定氧化锆、氧化钇		900		200
碳及其化合物				
LTI 热解碳 5%～12%硅		600		30
PAN AS4 纤维		3980	2.0	240
PEEK，61%碳纤维，长		2130	1.65	125
PEEK，61%碳纤维，±45		300	1.4	47
PEEK－30%碳纤维，切断		208	17.2	17
			1.3	
生物组织				
羟基磷灰石（HA）		100		114～130
骨		80～150	0.001	18～20
胶原		50	1.5	1.2

① F138 锻造不锈钢：17～19 铬，13～15.5 镍，2～3 钼，<2 锰，<0.08 或者<0.03 碳。
② F75 钴铬钼铸造合金：27～30 铬，<1.0 镍，5～7 钼，<1 锰。
③ F799 钴铬钼锻造合金：26～30 铬，<1.0 镍，5～7 钼，<1.0 锰，<1.5 铁，<1.5 碳。
④ F136 钛-6 铝-4 钒合金：5.5～6.5 铝，3.5～4.5 钒，<0.015 氮，<0.13 氧，<0.08 碳。



图 6.5　主要由金属制成的典型的全髋关节置换体

注：股骨柄上的球状股骨头上覆盖超高分子量聚乙烯（UHMWPE）衬垫，然后装入髋臼杯内，UHMWPE 衬垫使关节几乎可以无摩擦活动。照片所示是美国 Wright Medical Technology 公司的 PROFEMUR® Z 产品。

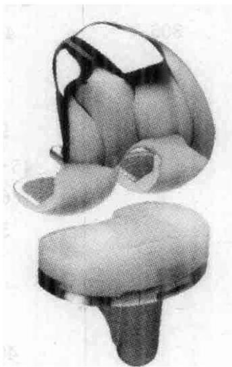
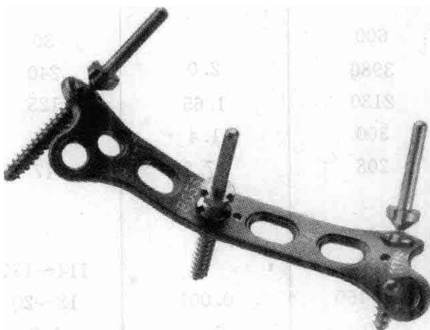
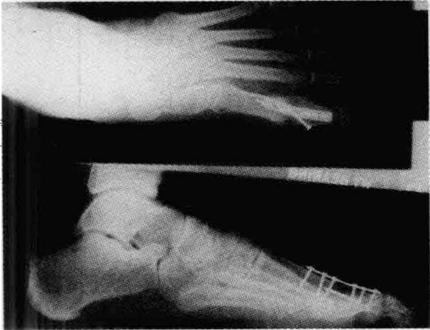


图 6.6　具有超高分子量聚乙烯（UHMWPE）承载表面的人工金属膝关节

注：照片所示是美国 Wright Medical Technology 公司的 ADVANCE®产品。



a)



b)

图 6.7　治疗过程中用于固定骨折部位的金属接骨板和螺钉（图 a），根据受伤程度，骨头痊愈之后，金属板和螺钉可能可以取出（照片所示是 NEWDEAL 商标的 HALLU®-FIX MTP 产品）；显示病人脚上和手上植入的金属骨板和螺钉的 X 射线图像（图 b）（此 X 射线照片由美国 Wright 医疗技术公司提供）

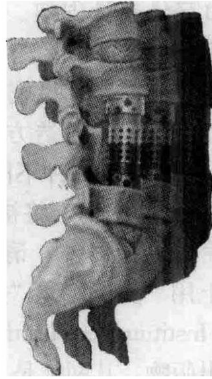


图 6.8 将脊椎骨连接起来的金属器械（即脊椎融合器），用于骨质疏松或背部受伤造成的脊椎骨损伤

注：这种金属笼可以根据病人自身的脊柱结构将相邻的脊椎骨跨接在一起（照片所示是美国 Medtronic Sofamor Danek 公司的 VERTESPAN®产品）。

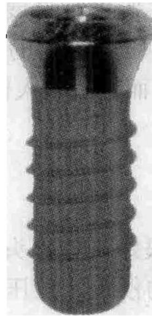


图 6.9 取代托牙的种植牙是用金属牙根植入体替代缺损的牙齿，然后再套上陶瓷牙冠

注：与托牙相比，种植牙的优点是其植入体可以将作用力传给颚骨，从而使颚骨能够继续接受刺激，避免逐渐衰退（照片由 Connecticut 大学保健中心的 Martin Freilich 博士馈赠）。

医疗器械材料的选择很复杂，需要考虑很多因素，包括力学承载的要求、材料本身的化学特性和结构特性以及各种生物需求。金属制成的膝关节和髋关节、接骨板和脊椎融合器等器械之所以能够长期使用，是由于它们具有很高的机械强度，并且具有较好的生物相容性。与陶瓷和聚合物等其他材料相比，金属的优点是坚韧并具有延展性；缺点是易腐蚀，这是由金属键（即自由电子）引起的。20 世纪初用于髋关节植入体的钢材料在体内腐蚀很快，曾经在康复过程中造成不良后果。因此，后来人们喜欢选用钛合金或者钴铬合金作为髋关节、膝关节和牙齿等植入体的材料。另外，与一般被替代的天然材料相比，金属材料具有较高的密度和很强的刚性，这也造成了一种应力遮挡的不利效应。金属关节植入之后，周围骨骼由于承受不到正常的作用力负荷会逐渐衰退，这就是所谓的应力遮挡。镍钛合金等一些特殊金属被称为记忆合金，它们在外力作用下会弯曲，但一旦外力撤除，又会恢复到原始形状。

这类金属已经被用于制作眼镜架和冠状动脉支架。这种动脉支架可以被压扁推入导管，一旦穿出导管，支架就会弹回原有的圆柱形。

金属器械的制作一般需要精密铸造、计算机辅助设计与制造（CAD/CAM）、研磨或粉末冶金等技术。某种特定医疗器械的制造方式取决于多种因素，如植入体的最终几何形状、金属的成形和制造特性以及各种不同制造方法的成本等。

例 6.2 假设某种支架是螺旋状的编织物，将被植入到堵塞的血管内用于增加血流量。这种血管支架必须是永久性的柔韧器件。请问什么材料可以满足此要求？除了利用本章所述内容以外，请在网上用“支架”和“金属”等关键词搜索最新的材料信息。美国国立卫生研究院（National Institutes of Health）的 PUBMED 网站收集了生物学、医学和生物医学工程等方面的学术出版物，其网址是 http://www.ncbi.nlm.nih.gov/entrez。许多公司网页也可以提供有关信息。其中，Guidant 和 Boston Scientific 就是生产冠状动脉支架的两家公司。美国专利局则是另一个很有用的网站，可以搜索各种材料在外科和医疗器械中的应用，其网址是 www.uspto.gov。

解： 制作血管支架的较为理想的材料是金属，如铂或钛，也可以用镍钛合金等形状记忆合金。用镍钛合金制作的支架具有自展性，也就是它们在体内展开时可以恢复成“记忆”的原有形状，特别适用于弯曲的血管或者粗细有变化的血管。这些金属材料具备血管支架所需的强度，而且植入人体后也已表现出很好的生物相容性。

■

6.2.3 陶瓷和玻璃

陶瓷材料的优点是生物相容性很好，与骨头的相容性尤其好，并且，它具备惰性，磨损率很小，可以抵抗微生物的侵入，抗压性很强。缺点是易碎，可能出现大破坏，且制造困难。这些特性由陶瓷的原子结构决定，它与金属不同。金属原子之间的金属键结合很松散，原子容易移动；而陶瓷的原子之间通过共价键结合成化合物形式，其原子之间不易移动，因此，陶瓷的导热和导电性能都很差。陶瓷的熔点和脆性这两个特性与金属的差别特别明显。陶瓷具有相当高的熔点，一般高于 1000℃，并且易碎。表 6.1 列出了医疗器械中常用的几种陶瓷材料。图 6.10 的照片则显示了髋关节植入体的陶瓷股骨头，图 6.11 所示是药片形状的硫酸钙骨移植片替代物（即人造骨）。

某些陶瓷、玻璃、玻璃-陶瓷等复合材料的组成成分对于骨头的连接具有直接的刺激作用，这种特性对于髋关节、膝关节和脊椎融合器等医用植入体的稳定性非常重要。这类材料就是所谓的生物活性陶瓷。对于体内取回的植入体的研究结果表明，在体内，植入体材料表面会形成具有生物活性的磷酸钙，这种生成的磷酸钙很像人体骨头的成分。材料表面一旦形成骨样矿物质表层之后，骨细胞就很容易紧密粘附于材料表面。因此，如果在植入体表面覆盖骨质磷酸钙材料，那么就可以达到同样的效果。

医用磷酸钙有多种不同的分子结构，包括羟基磷灰石、碳酸磷灰石、磷酸二钙



图 6.10　该人工髋关节的聚合物承载表面和某些金属部件被陶瓷取代，这样可以增加关节植入体的寿命，陶瓷的股骨头和髋臼杯是这种设计的特点（照片所示是美国 Wright Medical Technology 公司的 LINEAGE® 产品）



图 6.11　如果病人自身的或者他人捐献的骨头数量不足以填补缺损的骨头，那么可以使用磷酸钙或者硫酸钙制成的人造骨移植片（照片所示是美国 Wright Medical Technology 公司的 OSTEOSET® 产品）

或透钙磷石、β磷酸三钙或磷酸四钙以及无定形磷酸钙等。特定磷酸钙医疗器件的稳定性取决于其晶相、晶体大小及其完整性、处理时的温度、密度和使用环境等。在生理环境的温度和 pH 值下，羟基磷灰石是很稳定的晶相，一般经过很长时间才会产生生化分解而被吸收。不过，骨细胞以及巨噬细胞可以改变局部区域的 pH 值，使局

部环境呈酸性，然后启动细胞介导的吸收过程。可被生物体吸收的陶瓷一般属于非羟基磷灰石的磷酸钙盐，或者碳酸钙、硫酸钙等钙质生物材料。

陶瓷的高熔点使其不能用铸造的方法浇制成型，因此，陶瓷件一般用粉末制作。陶瓷粉末通过湿化学合成法或者通过原材料的研磨制备；然后，通过湿法或者干法制作成型。湿法成型是把陶瓷粉末加入含有粘合剂的溶液中调和成泥浆，再倒入模具制成生坯；干法成型则把陶瓷粉末干压成生坯。陶瓷生坯最后必须经过烧结，以便除去粉末颗粒之间的空隙，使陶瓷件致密。对于承载器械，必须几乎完全去除空隙，否则，残留的空隙会在器械内部形成微小裂纹，削弱其承载能力。而对于人造骨移植片等其他应用，则反而希望材料中含有骨骼的骨小梁或松质骨中存在的那种大孔隙，使细胞能够渗透到材料中并生长成新生骨。这些空隙一般用聚合物小珠等另一种物质相来形成。陶瓷生坯刚开始烧结时，其中混合的小珠占据一定的空间；到了烧结的最后阶段，小珠被烧掉，就形成了空隙。本章后面还会介绍多孔支架的制作方法。

玻璃的主要成分是二氧化硅。二氧化硅是一种网状结构的氧化物，加热后很容易达到熔点，因此，与陶瓷不同，它比较容易加工。

例 6.3 请问什么材料适合制作髋关节植入体的臼杯？在选材时需要考虑哪些因素？请用“陶瓷”、“髋关节置换”等关键词搜索各个科技网站、公司网站和专利网站，寻找相关信息。

解： 较早的髋关节臼杯由金属支撑结构和聚乙烯杯组成。但是，聚乙烯磨损之后会产生碎屑，为了解决这个问题，人们已经开发了氧化铝或氧化锆制作的陶瓷臼杯和股骨头的新产品。由于要与人造关节的头部直接接触，臼杯必须能够抗磨损、抗变形，且具有低摩擦的表面。在使用过程中，陶瓷材料之间摩擦产生的碎屑要比传统设计中金属与塑料之间摩擦产生的碎屑少得多。因此，理论上，与传统金属结合塑料的产品相比，陶瓷人工全髋关节的损坏几率要小得多。不过，与金属部件相比，陶瓷部件易碎且疏松性骨头不易承受，因此，如今两种髋关节置换体都在使用。

■

6.2.4 聚合物

聚合物的多样性使它很适合用作生物材料。例如，聚合物既可以软，也可以硬；有强度低的，也有强度高的；既可以抗蛋白粘附，也可以吸附蛋白；既可以具有生物降解性，也可以持久稳定；还可以用各种方法制作成复杂的形状。聚合物的缺点是：其强度要比金属和陶瓷小，会随时间产生变形，可能在灭菌消毒过程中遭受破坏，而且还可能在体内退化或损毁，或者释放出毒性物质。表 6.1 和表 6.2 列出了聚合物在医疗器械中的一些应用及其力学特性。

大分子聚合物由小分子通过聚合反应形成，聚合反应有加聚反应和缩聚反应两种。加聚反应通过单体反复加成产生，每次只加一个单体，从而链接形成高分子聚合物；而缩聚反应则是几个单体分子一起缩合生成高分子，这种反应会产生水等小

分子产物。聚合反应过程中一般还需要添加填充剂、增塑剂、稳定剂和着色剂等辅料，用于提高聚合物的化学、物理和力学性能。

聚合物可以分为热塑性或热固性两类。热塑性聚合物具有线性或分支结构，其固体形式就像意大利细面条，各条多聚链之间可以相对滑动。加热时，链与链之间的滑动就变得比较容易，于是，聚合物就会熔化并产生流动。因此，热塑性聚合物可以反复加热、熔化并成型。通过添加不同的配体可以获得具有不同特性的聚合物。例如，由于氯原子比较大，对分子之间的滑动有阻碍作用，因此，聚氯乙烯（即 PVC）比聚乙烯（即 PE）要硬。如表 6.2 所示，聚甲基丙烯酸甲酯（即 PMMA）与超高分子量聚乙烯（即 UHMWPE）相比，PMMA 的强度较高、更硬、更脆。其实，PMMA 是将乙烯单体的 4 个氢原子中的 2 个替换掉了，一个换成甲基（$-CH_3$），另外一个换成丙烯酸甲酯（$-COOCH_3$）。这些较大的侧基团使分子链之间较难滑动，于是就增加了材料的强度和弹性模量。同时，侧基的存在也使分子难以转向，难以形成规则有序的晶体结构。因此，这种非晶体结构的 PMMA 是透明的。

与热塑性聚合物相反，热固性聚合物的多聚链之间存在相互交联。加热时热固性材料不能熔化，而会降解。“热固”是指加热时会发生化学反应，从而形成三维空间上的交联结构。常见的“5 分钟环氧树脂”就是一种热固性聚合物。如果将该产品提供的催化剂和环氧树脂两种成分混合在一起，催化剂就会使环氧树脂分子链之间形成交联。一旦加工成型后，就不能再受热熔化。材料中形成的交联的多少决定了材料的力学特性，交联较少的环氧树脂可用于制作橡胶手套；多加些硫磺产生较多交联之后，环氧树脂可以制作汽车轮胎；再增加交联数量后，环氧树脂就可以做成汽车电池的坚硬外壳了。

水凝胶是一种交联结构的聚合物，吸水后它会膨胀，它在生物医学中的应用引起了广泛关注。制备水凝胶时，先通过辐射法或者化学法产生交联聚合物，然后将这种高分子网络链用水处理后就可以得到膨胀的水凝胶。最常用的水凝胶是聚羟基乙基甲基丙烯酸甲酯（Polyhydroxyethylmethacrylate，PHEMA），其含水量与生物体相似，并具有抗降解性，不会被人体吸收，可以承受加热灭菌处理，并且可以加工成各种不同形状。水凝胶可用于制作隐形眼镜、药物输送载体、创可贴、性器官再造材料、人工肾膜以及声带置换材料等。

制作聚合物医疗器械要用到很多技术，需要考虑多种因素。例如，使用热塑性材料还是热固性材料；如果用热塑性材料，那么材料软化的温度是多少，等等。热固性聚合物必须先制备成液态聚合物，然后倒入模子中凝固定型，定型后就不能再熔化。而热塑性聚合物可以多次反复定型，其定型方法有压制、注射、挤出等。用挤出法可以制成纤维或者薄膜，再经过拉伸或者滚压提高其强度等特性。组织工程中使用的多孔聚合物支架的制作方法是，先配备聚合物溶液，然后在溶液中加入非溶剂物质，将聚合物析出成型。如果在聚合物与溶剂的混合物中加入大小合适的微小颗粒物，那么，还可以在材料中形成可以让细胞渗入的大孔隙。去除溶剂之后，经过冲洗，就可以将微粒洗去，留下孔隙。本章后面还会介绍其他制作多孔聚合物

支架的方法。

例 6.4 什么材料适用于制作血袋、透析袋？需要考虑哪些因素？

解： 自 20 世纪 50 年代以来，聚氯乙烯（即 PVC）一直用于制作血袋。由于 PVC 本身是脆性材料，因此，需要添加邻苯二甲酸酯类增塑剂使其变软。但是，长时间使用时这种塑料袋中的增塑剂会逐渐溶解到袋内液体中。实验结果表明，大鼠食用大量增塑剂后，会诱发癌症；因此，人们正在开发其他塑料材料。透析袋是用低密度聚乙烯（即 LDPE）制成。选择这类医疗器械材料时需要遵循的原则是材料必须柔软，化学性质必须稳定，具有相当的惰性。 ■

6.2.5 天然材料

天然材料用有机物或者植物合成，它们的化学特性和结构一般都要比合成材料复杂。表 6.1 列出了目前医疗器械中常用的几种天然材料。蛋白质和多糖是天然聚合物，蛋白质中的定向键使其具有很强的力学特性。例如，尼龙是拉力最强的合成聚合物之一，而蚕丝的极限拉伸强度却比尼龙还要高。并且，蚕丝的弹性模量几乎是尼龙的 13 倍。陶瓷也有天然的，天然陶瓷一般都是钙质的，如磷酸钙骨骼晶体、碳酸钙珊瑚和海贝壳等。天然陶瓷的韧性一般比合成陶瓷强，不易断裂。这是由于天然陶瓷的微结构具有高度的组织性，可以防止裂纹的产生。材料中包含的小陶瓷晶粒具有精密的排列方式，并被有机基质材料分割成许多薄片，这样，材料中产生的裂纹只能被迫沿着弯曲的有机基质延伸。天然材料还包括供体组织，如骨骼和皮肤等。如果供体组织是病人自己身上的，则称为自体移植物；如果是他人身上的，则称为同种异体移植物；如果是牛或猪等其他物种的，则称为异种移植物。

与合成材料相比，天然材料具有较小的毒性和感染几率；不过，其生产或分离的费用一般较高。而且，各种天然材料之间的差别很大，很难保证材料的一致性，有时会影响材料的商品化及其推广应用。分离或提纯过程中一般要用溶剂提取所需成分，或者用溶剂除去不需要的成分，从而留下完好可用的天然材料。例如，这两种方法都可以用于胶原蛋白的制备。如果商品标签上注明胶原是可溶性的，那么，它就是从公鸡鸡冠等天然组织中通过胃蛋白酶处理提取出来的。纤维胶原是由肌腱等天然组织制备，通过盐溶液、脂溶液和酸溶液等处理步骤去除其中的非胶原蛋白和其他分子之后，剩余的就是胶原纤维。

生物聚合物也可以利用细菌生产。聚羟基丁酯（polyhydroxybutyrate，PHB）产品就是通过细菌发酵获得的。在精确的葡萄糖和丙酸配料的培养下，细菌会在细胞质中合成聚合物颗粒。将细胞破碎，经过冲洗和离心，就可以收集到聚合物，再烘干制成产品。这种聚合物的性质与聚丙烯和聚乙烯相似，但它可以降解成为人体内具有的天然成分。生物聚合物的这种理想环保特性使其迅速占据了几个非医用小产品的市场，如生物可降解单丝鱼网等。但是，由于生物聚合物的生产成本比石油化工聚合物的生产成本高 10 倍，因此，生物聚合物的推广和应用比较困难。

利用化学方法聚合天然单体也可以制备生物聚合物。虽然这种聚合物不是生物系统产生的，但它们由生物体的基本材料元素合成，具备生物相容性，无毒且生物可降解。其中乳酸聚合物近年来已经被广泛用于医疗器械，制作生物可降解缝合线、组织工程的支架等。血液和肌肉组织中都含有乳酸。另外，利用微生物发酵也可以将葡萄糖等己糖变成乳酸产品。聚乳酸类聚合物经常与聚乙醇酸类聚合物结合使用。

例 6.5　外科医生愿意选用什么材料来修复外科手术或者外伤造成的较大的骨骼损坏？需要考虑哪些因素？请在网上搜索生产骨骼修复产品的公司，并到美国矫形外科医生学会（American Academy of Orthopaedic Surgeons）和美国整形外科医生学会（American Society of Plastic Surgeons）的网站看一看。

解：自体移植骨是外科医生修复骨骼缺损的首选，这种组织具有活性，含有骨骼再生所需的活细胞和生长因子。同种异体移植骨（即脱钙的异体移植骨基质）是第二选择，与新鲜移植骨相比，脱钙骨基质的优点是柔韧性好，容易整合到骨骼缺损部位，且再吸收迅速，只要几个月就能够被吸收，而非脱钙移植骨却需要数年时间。并且，脱钙骨基质还会释放一种诱导骨骼形成的物质，即骨形态发生蛋白，这种生长因子最早是 M. R. Urist 博士于 20 世纪 50 年代发现的。■

6.2.6　复合材料

复合材料由两种或两种以上的成分组成。虽然单纯材料内部也会存在不同的亚单元结构，如不同颗粒或不同分子等，但这种情况不属于复合材料。复合材料是指包含两种或两种以上不同化学组分且相互之间由界面隔开的材料。例如，生物医学复合材料有骨骼愈合治疗中使用的碳纤维增强聚乙烯复合材料、羟基磷灰石颗粒增强聚乳酸复合材料等。材料中所包含的非连续相一般都比连续相坚硬且强度高，因此，称为“增强体”。

复合材料既可以通过两种成分的混合、浇铸、压缩制成，也可以通过化学反应将不同组分合在一起。例如，在纤维表面涂上一层聚合物或在纤芯中灌注聚合物，然后加热并加压，就可以制成高强度的复合材料。另外，通过化学反应将第二种物质相加到原始材料中也可以形成复合材料。制作中空柱形材料时，如果盘绕一根细丝，就可以制成高强度的器件。复合材料适用于制作需要同时满足多种性能的器械，如人工全关节、牙科填料和接骨板等。复合材料的优势是几乎可以根据任何应用需求来设计其性能；然而，要制成一种理想的复合材料结构其实很难。其中主要的问题是，在两种物质相之间，第二物质相的扩散或者较弱的表面键合会导致材料力学性能的削弱，从而降低产品的性能。不过，在很多情况下，复合材料的实际性能仍然要比单一组分的生物材料要好，因此，它在生物医学上的应用范围正在不断扩大。

例 6.6　牙齿修复最好选用什么材料？复合材料与单一成分材料相比有哪些优点？

解：由汞、银和锡组成的银汞合金是最常用的牙科填料；另外，还有美观舒适

的有色牙科填料，它们用树脂制成，比如用双酚 A-甲基丙烯酸缩水甘油酯（即 BisG-MA）大分子加上二氧化硅微米或纳米微粒制成的牙科填料。聚甲基丙烯酸甲酯（PMMA）是制作全牙托和局部牙托的主要材料。铬合金则用于制作活动牙托的框架。牙冠和牙桥用金属支架外套齿色陶瓷制成。现在也有全陶瓷牙齿。最近，有人还开发了用玻璃纤维强化的光固化树脂复合材料，用于制作牙桥，并正在用这种复合材料替代以前用于支撑牙冠的金属牙桩。这主要是为了美观，这种聚合物牙桩不会像金属牙桩那样从牙冠下露出来。 ■

6.3 生物材料设计和选择中的仿生学

6.3.1 天然组织的基本构建机制

生物医学工程师负责设计各种医疗器械和系统，用于人体功能的修复、监测或辅助。仿生学是模仿或复制天然生物机理的科学，它常常是医疗器械成功的关键。天然组织和器官具有难以想象的复杂机制，至今仍然远远超出科学家们的复制能力；而且，组织和器官的各种精确功能还没有被彻底弄清楚。由于这些原因，从理论上设计医疗器械非常困难，医疗器械的设计总是要经过多次试验和失败才能成功。不过，通过对各种生物组织结构的研究，人们已经总结出如下几条基本原理，作为致力于组织和器官再生的生物材料科学家们的设计策略和指南。

1）细胞按照人体的基因编码生长并分化成为各种组织和器官；

2）细胞能够生产蛋白质、多糖、糖蛋白类和脂质等物质，并自行装配成具有多种形式的复合型胞外基质，支持生物组织的生长；

3）细胞之间通过生长因子等物质实现信息传导与交流，细胞的命运就是由这些蛋白质信号分子决定的；

4）血管在组织生长过程中具有至关重要的作用，它能够提供营养物质，带走废物，并且能够为组织的进一步生长提供细胞；

5）神经系统负责人体所有功能的整合和控制；

6）骨骼组织在蛋白质控制的成核过程中不断生长，它含有胶原蛋白微环境下生长出来的离散的纳米矿物质微粒晶体，因此骨骼很坚硬。

总之，天然组织的结构层次很分明，一个结构中含有另一个结构，就像鸡蛋的嵌套结构，或者像树枝的一层层分叉结构那样，相同形式的主导结构会在多个不同尺度上重复，使生物组织具有特定的强度和功效。仿生学的研究基础就是生物体基本结构知识和发育生物学理论，其研究结果为生物材料的设计和制造提供了合理的起点，尤其是对于组织工程化的再生型医疗器械。

6.3.2 天然组织的细胞构成

人体内含有的细胞种类超过 200 种，大约有 100 万亿个细胞。骨组织就有 10 种

以上细胞，它们是：

1）促使骨骼生长的成骨细胞；

2）分解骨骼的破骨细胞；

3）维持骨骼的骨细胞；

4）骨骼中血管所含有的各种细胞，如红血球以及淋巴细胞、单核细胞、巨噬细胞、嗜中性粒细胞和嗜酸性粒细胞等免疫细胞；

5）骨髓中所包含的脂质细胞以及尚未分化成特定成熟细胞的骨髓间充质干细胞。

我们人体中存在一类可以分化成若干种组织细胞的细胞群体，这就是干细胞，它们是人体修复受损组织的关键因素。即便是成年人，人体仍然具备利用干细胞生成新组织的能力。本书第 7 章的组织工程将详细介绍干细胞。干细胞可以通过响应各种刺激信号，分化成为特定的细胞种类。这些信号可以来自相邻细胞，也可以来自周围环境的化学和物理刺激，或者来自细胞外基质传导过来的机械作用力等。

细胞不仅是组织结构的构建要素，而且它们还会维护组织，使组织结构能够适应环境的变化，包括机械负荷的变化等。各种组织形成过程中的关键因素有：

1）细胞与细胞之间的信号传导（cell-cell signaling）。细胞之间的相互作用受控于细胞间复杂的分子通信。细胞的信号分子一般都是蛋白质，包括细胞因子和趋化因子。其中，细胞因子也称为生长因子，它们会引起细胞的增殖（即细胞复制）和分化；趋化因子则会诱导细胞迁移。细胞之间可以通过各种不同的连接方式互相粘附在一起。一方面，这些连接能够进一步增强组织的机械强度；另一方面，细胞之间的半透膜屏障也可以调节体液、离子和蛋白质的输运。

2）细胞凋亡（Apoptosis）。作为组织生长和形态形成的协调功能之一，某些细胞会在基因控制下自主性死亡，被称为细胞凋亡。这说明生物事件的启动是受控制的，而其终止也具有一定的机制。这是生物材料工程师在设计工作中可以利用的机理之一。细胞凋亡以非常整洁的方式进行，无用的细胞逐渐萎缩、变小，然后分裂成由细胞膜包裹的小颗粒，再由巨噬细胞清除干净。

3）细胞坏死（Necrosis）。与细胞凋亡不同，由创伤或者中毒引起的细胞死亡称为细胞坏死。细胞坏死过程中，细胞产生膨胀并爆裂，释放出内容物。由于细胞内含有信号蛋白质，细胞坏死会导致不利且失控的细胞活动。所以，细胞凋亡才是正常生物组织的细胞死亡模式。

4）血管生成（Angiogenesis）。如果没有氧气和营养的供应，或者缺乏废物清除的途径，那么细胞就不能存活。血管系统就是提供这些功能的器官，淋巴系统则能够协助清除多余的废物。组织在生长过程中必须迅速形成新血管，否则新生组织就会坏死。邻近的巨噬细胞、血小板或细胞外基质都会释放血管生成因子。

5）神经发生（Neurogenesis）。神经系统由几种细胞组成，包括传导电脉冲的神经元，保护、支持并滋养神经元的胶质细胞，以及构建脂质绝缘髓鞘的许旺氏细胞。正如本书第 3 章所述，神经系统负责人体各种功能的整合和控制。新生成的组织必须与神经系统建立连接之后才能工作。

例 6.7 血管生成是否总是必要且有利的？

解：在某些情况下，如新组织的生长过程中，血管生成是必需的。但是，有时血管却是不利因素，因为它们可能会带来过多的免疫细胞，导致机体自身免疫系统疾病，或者可能给肿瘤细胞提供营养，从而促进肿瘤生长。为了抑制这类有害血管的生长，制药厂商已经开发了抗血管生成因子及其生物材料载体。■

6.3.3 细胞外基质——天然生物材料支架

细胞外基质（Extra Cellular Matrix，ECM）是细胞的物理化学支持环境。从生物材料学上看，它是纤维蛋白支撑结构和胶状物组成的一种复杂的网状聚合物材料，其中的蛋白聚糖吸水之后会膨胀，产生一定的抗压强度。本书第 7 章将详细介绍 ECM 的成分及其功能。实质上，ECM 是细胞的微环境，它给细胞提供了吸附表面和迁移表面，以及蛋白聚糖的粘附位点，还提供了细胞信号蛋白分子释放的调控途径。它是一种从物理上和力学上将所有细胞连接在一起的支架。细胞通过特异性细胞表面受体粘附在 ECM 上，这些受体能够控制细胞的增殖、分化以及蛋白质表达。由此可见，细胞与 ECM 之间的相互作用在组织再生过程中具有非常重要的意义。

ECM 由细胞分泌出来的蛋白质自发折叠构成，主要是各种胶原蛋白（有 10 余种）以及纤粘连蛋白、层粘连蛋白、骨桥蛋白等大分子糖蛋白。ECM 是动态变化的，它不断受到细胞的调节。蛋白质的自发性肽链折叠被称为自组装。胶原蛋白分子由 3 条 α 肽链经过自组装形成相互缠绕的 3 条螺旋链。这 3 条肽链的氨基酸组成中每隔两个氨基酸残基就出现一个甘氨酸，它们与肽链中的脯氨酸残基和羟脯氨酸残基一起构成了稳定的三联螺旋结构。然后，这些胶原分子再进一步聚集，每几个分子平行排列成一组，相互重叠 1/4，形成交错排列的直线束，最后，这些直线束再聚集排列成胶原纤维束（见图 6.12）。各个胶原蛋白分子末端的基团之间留有空隙，为



图 6.12 骨骼各个层次上的结构

注：骨骼中至少有如下 5 层微观结构：股骨骨骼；柱状骨单位中的骨板，它们是皮质骨的基本组成结构；由微原纤维组成的胶原纤维；微原纤维上各个分子末端之间存在规则的间隙，这些间隙是骨的矿化成核位点；胶原蛋白分子。

以后的矿化留下空间。本章后面会进一步举例说明骨骼矿化的生物学机制。DNA 分子的组装有点像胶原蛋白，它是两条互补 DNA 链缠绕在一起形成双螺旋结构，然后这些螺旋状物质再进一步聚合。自组装是生物组织构建的关键机理之一，它使生物组织在分子水平上产生高度有序的结构。生物材料领域的科学家们目前正在利用这种原理制造纳米尺度上的复杂微型结构。

6.3.4 天然组织的层次化结构

天然组织具有像鸡蛋或者俄罗斯套娃玩具一样的层次化嵌套结构，当你打开外层时总可以看到里面有更小的一层结构，相似的结构形式在不同尺度下不断重复，生物组织就是这样组装成型，形成自身的特性和功能。在细胞介导的组织构建过程中，最小的组成单元首先实现自组装，然后这些单元再组装成较大的单元，最终完成更大单元的自组装。人体的功能性组织或器官都是这样构成的。组织中的细胞一般都包埋在一层层细胞外基质中，细胞的生存和活动维护着整个组织的环境。例如，骨骼在纳米、微米和宏观层次上都具有相似的纤维状结构。在骨骼最小的层次上，α 肽链组装成三联螺旋结构的 I 型胶原蛋白分子，然后这种胶原分子组装成微原纤维，微原纤维再组装成更大的排列成薄板状的纤维，薄板会像夹板一样一圈圈环绕血管形成具有管状或大纤维状的骨单位（见图 6.12）。骨骼肌也有层次结构。肌动蛋白丝和肌球蛋白丝构成肌小节的直线形结构，然后聚集成肌原纤维束，再聚集成肌纤维，这种肌纤维就是含有上百个细胞核的肌细胞，最后多个肌纤维形成所谓的肌肉。本书第 3 章有肌肉结构的示意图。

染色体的生物结构同样具有层次化。在自然效应作用下，在不同长度范围内经过多次不同尺度上的弯曲和缠绕，7cm 长的 DNA 链竟然能够缩短到原长度的 1/10000，从而形成染色体。在纳米水平上，DNA 是单股双螺旋结构；然后，每段长约 146 碱基对的 DNA 双链盘绕在组蛋白上形成一个个核小体，核小体彼此连成串珠状的染色质细丝；染色质细丝再缠绕成更短更粗的纤维；这种较粗的纤维进一步盘绕后再缩短成为染色单体，两条染色单体最终连接成一条染色体。

6.3.5 生物组织的矿化

骨头、鹿角、牙齿、珊瑚、蛋壳和贝壳都是生物组织矿化的例子。虽然这些组织的功能各不相同，而且外形也有很大差别，但是，它们都由大量很小的磷酸钙或者碳酸钙离散晶体颗粒与蛋白基质结合形成。矿质晶体的成核与生长过程受到有机蛋白成分的调控，细胞分泌的有机蛋白分子会通过自组装构成一个模板，给矿物质的生长和成核提供规则的空间。以下 3 条有关各种矿化组织的合成、构建和组装的基本生物学原理对于材料科学家和工程师的工作具有非常重要的指导意义。

1）生物矿化在特定微环境的亚单元池中进行，这意味着某些功能位点受到诱发晶体形成的刺激信号的作用，而其他某些位点则受到抑制结晶过程的信号的作用。

2）某种特定的矿化相具有一定的晶体大小（通常在纳米尺度范围）、形状和生长方向。

3）矿化组织的宏观结构是通过很多逐渐增大的单元组装形成，从而构成层次化的独特的网状复合材料，使材料拥有独特的性能。

生物矿化还有一个重要特点：在大多数组织中，如果需要优化组织的强度、调节机体的生长、维持矿物质的离子平衡状态，或者修复组织，那么原始的矿物质结构就会产生重塑过程。

骨骼中，受到调控的矿物质的生长和成核过程在胶原蛋白基质构成的微池中进行。骨单位分泌的Ⅰ型胶原蛋白分子经过自组装后形成微原纤维，微原纤维具有独特的结构，它由周期性重复的长约67nm的胶原分子与40nm的间隙排列而成（见图6.12）。这些间隙构成了微环境，其中含有游离的矿物质离子以及附着在胶原分子上的磷蛋白的侧链基团。周期性出现的分子之间的间隙形成了矿化相成核的场所，于是在胶原纤维这些离散的空间上就形成了薄板状的磷灰石晶体，而且，在这种独特的空间限制下，矿化晶体不可能自由生长，它们被迫成为离散的非连续晶体。在骨骼的正常矿化过程中，即碳酸磷灰石的生成过程中，只有一种磷酸钙相会结晶成核，并且矿物质沿着特定的结晶方向生长。由此可见，这种骨组织的形成过程包含了天然生物组织构建时的所有关键要素：细胞、细胞和矿物质的胞外基质微环境、蛋白质信号以及层次化的结构。

6.4 材料与生物组织之间的相互作用

6.4.1 材料与血液和蛋白质之间的相互作用

生物材料的植入往往会造成创伤，会引起出血，因此，血液一般会首先与植入材料相接触。血液是一种混合物，其中包含水、各种细胞和细胞碎片（即血小板）、盐类以及蛋白质（即血浆）。蛋白质是生命体的主要组成成分，它们由20个氨基酸组成的长长的链肽通过肽键连接而成。人体内的蛋白质几乎具有无数的功能，它们既可以作为酶催化数千种维持生命的重要化学反应，也可以成为负责细胞迁移和增殖的细胞信号分子。此外，蛋白质还是许多生物组织细胞外基质的基本结构材料。蛋白质及其结构的改变会导致蛋白质功能的变化，从而引起人体产生疾病。因此，血液中某些蛋白质成分的检查可以反应某些疾病或癌症的状况。

蛋白质在决定生物组织与植入体之间界面的最终性质上具有重要的作用。生物材料可以制作成能够选择性地吸附蛋白质，从而促进细胞和组织在材料上的粘附和活动；也可以制作成能够排斥蛋白质，从而抑制组织与材料之间的相互作用。值得注意的是，生物材料的植入会使生物体微环境中的pH值和离子强度等因素发生变化，从而改变附近蛋白质的构象，进而改变其功能。此外，蛋白质在与生物材料的

固体表面发生作用时，也会产生结构变化，从而失去其具有的某些生物活性。例如，血液中含量最多的蛋白是白蛋白，其次是保护免疫系统的免疫球蛋白；但是，由于被吸附的蛋白质可以发生置换，因此，最终吸附在材料表面上的蛋白层可能并不是这两种蛋白，而是纤维蛋白原。虽然纤维蛋白原在血液中的含量并不高，但是它对生物材料表面的亲和力却较强。

血液中的蛋白凝血因子 XII 吸附在植入的生物材料表面之后，会导致血液的凝固。这种凝血因子被材料吸附后，血液中的血小板就会跟着吸附到材料上，从而导致纤维蛋白凝块的形成。由此可见，初始的蛋白吸附会引起一系列细胞的级联反应。血液与材料之间接触所引起的生物作用会产生各种细胞和细胞因子，因此，每种与血液接触的生物材料都会诱发人体的生物反应。

6.4.2　生物材料植入后创口的愈合反应

生物材料植入时会造成生物组织的断裂，形成创口，并立即由此触发机体高度发达的创伤愈合反应，本书第 7 章将介绍创伤引起的细胞正常响应过程，这些知识是了解和预测组织与材料之间相互作用的基础，这些相互作用总是包含如下 4 个阶段（见图 6.13），它们在时间上会相互交叠：

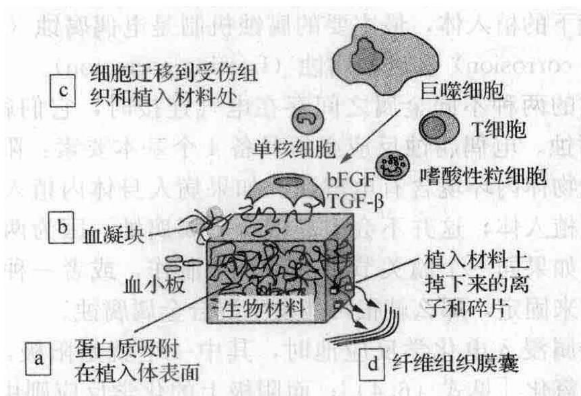


图 6.13　组织与生物材料之间的正常相互作用一般包括在时间上相互交叠且相互依赖的 4 个伤口愈合阶段：止血、炎症、增殖与初始修复以及组织重塑

注：图中 a 为蛋白质在生物材料表面的吸附导致各种细胞作用的发生；b 为血凝块的形成完成止血；c 为血液中包含的细胞以及其他炎性细胞努力处理异物（即植入体材料）并修复其周围的组织；d 为生物机体利用纤维组织将植入材料包囊起来，从而保护自身免受侵扰。

1）止血：在此阶段，血小板细胞通过与生物材料表面附着的蛋白质形成粘连并释放凝血因子，控制出血。凝固的血块形成组织修复的初始基质，填补了植入物周围的空隙。

2）炎症：血凝块的形成会诱发细胞信号分子（即细胞因子）的产生，细胞因子会将附近血流中的炎性细胞聚集起来。炎性细胞包括嗜中性粒细胞、单核细胞、淋巴细胞和巨噬细胞。它们会通过所谓的吞噬作用，尽可能消化组织碎片和生物材料。

同时，炎性细胞在伤口处释放的生长因子开始启动伤口边缘处结缔组织细胞的分裂（即细胞复制）。

3）增殖与初始修复：在各种炎性细胞所释放的生长因子的作用下，生物材料周围会发生细胞增殖和聚集，从而再生受损组织。于是，位于创伤中心的没被降解的生物材料一般就被纤维组织严密地包裹起来。这层纤维膜囊把材料与其周围生物环境隔离开来，从而保护了生物机体。炎性异物反应的强烈程度决定了纤维膜的厚度。影响纤维膜厚度的其他因素还有植入物的化学特性、形状和物理特性，植入物释放药物或腐蚀物质的速率、聚集程度及其生物活性等。如果是永久性植入，植入物不发生生物降解，那么小膜囊就会一直伴随着植入体。只有骨骼植入物没有膜囊包裹，骨质可以直接附着于磷酸钙植入体的表面，不会生成纤维组织层。

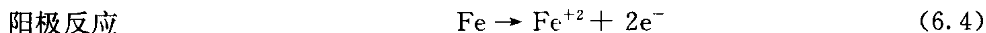
4）组织重塑：迅速形成的新组织会被细胞重新塑造成与原始组织比较相近的功能性组织，但一般会留下疤痕。

6.4.3 金属腐蚀

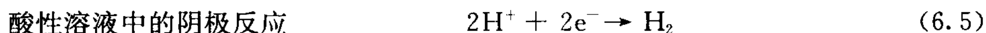
金属腐蚀的机制有好几种，抗腐蚀是金属植入物必须解决的重要问题之一。处于各种盐溶液环境下的植入体，最主要的腐蚀机制是电偶腐蚀（galvanic corrosion）、缝隙腐蚀（crevice corrosion）和磨损腐蚀（fretting corrosion）。

当浸入电解质的两种不同金属之间存在电气连接时，它们就会发生电偶腐蚀，也称为混合金属腐蚀。电偶腐蚀反应必须具备4个基本要素：阳极、阴极、电解质和外电路导体。生物体内环境含有电解质。如果病人身体内植入了不同合金制作的左右两个全髋关节植入体，这并不会引起混合金属腐蚀，因为两个髋关节之间没有电气连接。但是，如果同一个髋关节用两种合金制作，或者一种金属制作的接骨板用另一种金属螺钉来固定，那么就很可能发生混合金属腐蚀。

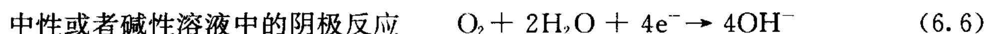
当两种不同金属浸入电化学反应池时，其中一个就是阳极，另一个则是阴极。阳极上会发生金属氧化［见式（6.4）］；而阴极上的化学反应则由环境的pH值决定［见式（6.5）和式（6.6）］。如表6.3所示，反应的方向由电动势（即电压）的大小决定。表中所列电压值表示1mol/L的离子浓度下，反应平衡时金属的半电池电位，其中氢的电动势定义为零。由表中可见，铁的标准电动势是－0.44V。如果铁与电动势为0.34V的铜连在一起，那么，两者之间的电位差就是0.78V。其中，铁是阳极，会发生式（6.4）所示的氧化反应；铜是阴极，根据溶液pH值的不同，阴极上会发生式（6.5）或者式（6.6）所示的反应。



阳极反应 $$Fe \rightarrow Fe^{+2} + 2e^- \tag{6.4}$$



酸性溶液中的阴极反应 $$2H^+ + 2e^- \rightarrow H_2 \tag{6.5}$$



中性或者碱性溶液中的阴极反应 $$O_2 + 2H_2O + 4e^- \rightarrow 4OH^- \tag{6.6}$$

表 6.3 金属电动势大小的排序表（表中所示是 25℃水溶液中的标准还原电位）

金 属	标准还原电位/V	
$K = K^{+} + e^{-}$	−2.93	↑ 活性增强（即阳极化）
$Na = Na^{+} + e^{-}$	−2.71	
$Al = Al^{+3} + 3e^{-}$	−1.66	
$Ti = Ti^{+2} + 2e^{-}$	−1.63	
$Zn = Zn^{+2} + 2e^{-}$	−0.76	
$Cr = Cr^{+3} + 3e^{-}$	−0.74	
$Fe = Fe^{+2} + 2e^{-}$	−0.44	
$Co = Co^{+2} + 2e^{-}$	−0.28	
$Ni = Ni^{+2} + 2e^{-}$	−0.25	
$Sn = Sn^{+2} + 2e^{-}$	−0.14	
$H_2 = 2H^{+} + 2e^{-}$	0.000	
$Cu = Cu^{+2} + 2e^{-}$	+0.34	
$Ag = Ag^{+} + e^{-}$	+0.80	
$Pt = Pt^{+2} + 2e^{-}$	+1.20	惰性增强（即阴极化） ↓
$Au = Au^{+3} + 3e^{-}$	+1.50	

由于单位摩尔溶液的自由能由其浓度决定，因此，电化学反应池的自由能变化和电极电位由电解液的组成决定。于是，反应的方向和速率也取决于溶液的浓度。在上述铁与铜组成的反应电池中，如果增加溶液中的 Fe^{+2} 离子浓度，铁的电动势就会向正方向偏移（即趋于惰性化）；于是，铁与铜之间的电位差就会减小。同理，阴极处的氧浓度也会影响其电动势的大小。增加氧浓度会使其电动势变得更正（即惰性化），而降低氧浓度则会使其电动势变得更负（即活性增加）。实际上，下面将介绍的缝隙腐蚀就是由氧浓度的变化引起的。

上述原电池反应不仅会发生在不同合金之间，还会发生在同一块合金内部。碳化物等颗粒物的各种界面上以及合金内部不同相之间也存在不同的电动势，它们都构成了原电池。金属的冷处理加工还会增加其自由能，使其更容易腐蚀。例如，弯曲金属板或者敲打铁钉头部，都是局部冷处理，会使这些区域相对于其他部分呈现阳极化。

不过，利用电偶腐蚀的原理，通过阴极化的方法可以保护某些部分免遭腐蚀。例如，在钢铁船体上粘贴锌块就可以防止船身生锈，这些锌块其实是作为阳极牺牲物保护了钢铁船身。锌块同时可以保护船上的金属泵等其他金属部件。另外，如果在地下钢铁管道上的某个位置连上电压源，使管道成为阴极；而将另一个容易替换的部件作为阳极，那么就可以保护管道免受腐蚀。

电极的大小也会影响电化学反应的速率，镀锌铁与镀锡铁之间的区别就是一个典型的例子。如表 6.3 所示，锌比铁的活性高，它与铁组合时是阳极；因此，铁的表面镀上锌之后就是加上了一层阳极材料。如果有小部分锌的镀层被刮擦，露出了铁，那么由于阴极铁的暴露面积很小，反应很有限，腐蚀就很小。相反，锡相对于铁是

阴极。如果锡镀层被刮擦，露出的小面积铁阳极与大面积锡阴极形成电化学反应，刮擦部位阳极腐蚀造成的铁损失就会导致铁的暴露面积不断扩大，腐蚀速率也随之增加。如果在锡阴极上涂上涂料或清漆等非导电材料，那么就可以防止这种自我加速的腐蚀过程的发生。

暴露在氯化物溶液中的狭窄空间中会发生所谓的缝隙腐蚀。如金属与非金属之间的垫圈式连接、用螺钉或者夹具固定在一起的两块金属之间等地方都会存在这种狭窄的空间。缝隙腐蚀需要经历数个发展过程，可能需要半年至 2 年，才能形成具有一定浓度的电化学反应池。一些由多个金属部件连接在一起的植入器械会发生缝隙腐蚀，如某些全髋关节置换件、用于固定骨折的接骨板和螺钉以及某些矫正器具等。

缝隙腐蚀的初始阶段，腐蚀在缝隙内部和外表面上均匀发生；也就是到处都存在阳极和阴极反应，阳极的金属被氧化，阴极的氧气和氢氧化产物则被还原。经过一段时间之后，由于缺乏大分子氧气的流通，缝隙中的氧气几乎耗竭，于是，缝隙内阴极的氧气还原反应就会停止，但其中的金属氧化反应仍然继续。金属氧化释放的电子会通过金属传导到缝隙外的游离表面上，并在那里被还原反应消耗。这会导致缝隙中过量正电荷的聚集，使得带负电荷的氯离子流入缝隙，并与其中的正离子中和。金属氯化物在水中发生水解时会产生不溶于水的金属氢氧化物和游离酸［见式（6.7）］。这将导致缝隙中的酸浓度不断增加，从而形成自我加速反应。



$$M^+Cl^- + H_2O \rightarrow MOH + H^+Cl^- \tag{6.7}$$

如今外科手术使用的合金都涂有一层稳定的氧化膜，正是这层保护膜使各种合金具有抗腐蚀性，这种膜称为钝化膜。例如，表 6.3 中的钛属于一种活性较强的金属，它可以形成粘性很强的氧化物，防止腐蚀。各种不锈钢和钴合金器械上可以形成铬氧化膜。如表 6.3 所示，这些金属的活性都较高，但形成氧化膜之后就成为惰性了。

不锈钢要实现自我钝化，必须至少含有 12%的铬。碳对于铬具有很强的亲和力，碳化铬的平均分子式是 $Cr_{23}C_6$。不锈钢合金块中的铬原子会移向其中的碳化物，从而形成碳化铬。这使得碳化物具有较高的铬含量，而合金中其他部分的铬却不断减少。如果合金中的铬含量低于 12%，那么就会影响以后的再钝化，不锈钢就容易受到腐蚀。因此，为了安全，外科医用不锈钢一般都含铬 17%～19%，而含碳量则控制在低于 0.08%或低于 0.03%的水平。

对于不锈钢焊接部件，尤其要注意碳化物形成的问题。如果钢材被加热到 425～870℃的温度范围内，铬就会在材料中扩散并形成碳化物。如果温度超过 870℃，碳原子晶格就会熔化。如果温度低于 425℃，那么，因铬的流动性太小，就不会形成碳化物。因此，如果焊缝之外金属的最高温度在 425～870℃范围内，那么碳化物就会形成，这就是所谓的金属焊缝两侧的焊接腐蚀。焊接之后如果进行热处理，可以使碳化物重新溶解，然后再迅速淬火，就可以避免碳化物的再形成。

当氧化膜完好无损时，外科医用合金材料表现出惰性。但是，一旦氧化膜破损，如刮伤或磨损，暴露的金属就具有活性。这时，再钝化过程可以恢复惰性。磨损腐

蚀就是指保护膜的不断磨损以及由此引起的暴露金属的不断氧化过程。例如，遭受缝隙腐蚀的器械内存在的磨损腐蚀会加速缝隙腐蚀。

例 6.8　美国北部地区的旧汽车往往在车门底部和行李箱盖的底部会生锈，并且无盖货车的后挡板也常常生锈。请列出发生这种情况的 5 个原因。如果在这些部件上拧上锌块，会有用吗？

解：1）车门和行李箱盖的边缘和底部都是通过金属的弯曲形成，弯曲时的冷处理使得受弯曲部位相对于其他部分成为阳极化。

2）卷曲部分通过点焊来封闭，这会造成不同的微结构，从而形成电偶腐蚀。

3）冬天，为了防止冰冻，美国北部地区公路上常会撒上盐，飞溅的盐水会滞留在卷曲处，那就是电解质。泥沙也会滞留在卷曲处，有利于维持潮湿的环境。

4）汽车生产厂家在行李箱盖的底边装上镀铬装饰钢条。铬相对于钢是阴极，于是，就又形成了一种电偶腐蚀。

5）公路崎岖不平，会导致汽车行驶时行李箱盖不停地蹦跳，箱盖上的油漆会被擦掉，暴露出金属，从而产生磨损腐蚀。

只有当金属腐蚀发生在同一个电解池中时，加上锌块才会有用。因此，在缝隙中加上锌片可以缓减缝隙腐蚀。如果汽车掉进大海，那么锌块就可以保护整个汽车了。 ■

例 6.9　假如你祖母身上已经安装了一个不锈钢全髋关节。现在，她要把另一个髋关节也替换掉，医生想用钴铬合金制成的髋关节。这会有腐蚀问题吗？

解：不会。因为 4 个导致电偶腐蚀的基本要素之一不存在。不锈钢髋关节是阳极，钴合金是阴极，人体的盐溶液是电解质，这 3 个要素都具备了；但是两个髋关节之间缺少电气连接，所以不会有腐蚀问题。如果她跌倒并造成了骨盆骨折，骨折部位加了外部固定器，那么，就可能形成电气连接，从而产生腐蚀问题。不过，这些合金非常耐腐蚀，并且具有类似的电化学特性，因此也许不必担心。 ■

6.4.4　生物材料的降解和重吸收

生物材料可以是永久稳定的，也可以是可降解的。降解过程可能通过化学反应进行，也可能由细胞完成。生物可吸收植入体在体内环境中会随时间的推移逐渐发生降解，并被天然组织所取代，其使用目的是在天然组织的再生过程中暂时提供足够的支持强度和细胞供给。生物材料化学性质和结构的微小变动都可能改变其吸收速率，这个特点可以用于设计各种不同用途的材料，不过，它也可能导致器械的意外损坏。胶原蛋白、聚乳酸（PLLA）、聚乙醇酸（PGA）以及聚乳酸和聚乙醇酸的共聚物（PLGA）等都是常用的可吸收材料。PLLA 和 PGA 通过聚酯键的水解过程降解。小分子量的植入物质的分解会产生小碎片，从而诱发巨噬细胞的免疫反应。PLLA 和 PGA 的降解时间大约在半年至数年不等，由材料的原始分子量和结晶程度决定。PLLA 和 PGA 两者的共聚物一般在几个月内就会降解成碎片，这些乳酸和乙醇酸碎片最终代谢成为二氧化碳和水。磷酸三钙陶瓷材料通过表面溶解过程会降解

并变成钙盐和磷盐，而钙盐和磷盐都是人体的天然成分。

生物材料的降解可能导致伤口愈合停留在某个阶段，从而形成未能愈合的慢性创伤。如果材料降解过快，并释放颗粒物拖延炎症期，就会发生这种情况。持久的炎症会导致多核巨细胞的形成，这些细胞会不断清除侵入物，它们是异物反应的标志。这种持续的炎症很可能最终不得不导致外科手术取走植入物。另外，即使愈合过程能够顺利地进行到纤维膜囊形成阶段，仍有可能产生问题。例如，药物输送植入体的降解作用所导致的纤维膜囊的形成，会影响药物的释放，最终使药物输送植入体不能正常发挥作用。

例 6.10 顺利通过伤口愈合的所有 4 个阶段之后，是否仍然可能产生生物材料的降解，并导致已经愈合的伤口重新恶化，为什么?

解: 这是可能的。某些接受全关节置换的病人已经出现过这种情况。人工髋关节等全关节置换件一般都由两个金属部件组成，中间夹一层聚合物制作的承载表面(通常是超高分子量聚乙烯，即 UHMWPE)。关节弯曲时，金属表面与较软的聚合物表面会摩擦并产生碎屑，从而导致巨噬细胞的会聚。这些细胞识别碎屑颗粒物，将其视作异物清除掉。由于这些合成材料颗粒并不能被细胞含有的酶降解，因此，炎症会无休止地持续下去。过量的炎性细胞因子会不断破坏植入体附近新愈合的骨骼，导致植入体的松动。幸运的是，这种聚乙烯材料已经有了很大改进，如今的髋关节置换体不再像早期产品那样容易产生磨损屑碎。同时，人们还开发了陶瓷连接陶瓷的髋关节，它们有更好的耐磨性，且耐腐蚀。不过，陶瓷人工髋关节不适合老年人骨质疏松的骨骼，因为这种材料非常硬。 ■

6.4.5 免疫原性

免疫原性是指物质诱发免疫反应的特性。细菌、草木的花粉、生物材料的小颗粒或可吸收颗粒以及可以导致过敏或发炎的某些食物蛋白等都是免疫原。实质上，保护我们人体的整个免疫系统包括皮肤等构成的物理屏障、酶和抗体等组成的化学屏障以及 T 细胞等形成的细胞屏障。生物材料植入体内之后，免疫系统中的相关蛋白质就会立即附着到植入体表面，从而引导随后的细胞行为。这里必须再次强调，生物材料表面的化学性质和结构在决定蛋白质附着物的种类和附着程度上具有重要作用，它们决定了组织与生物材料之间的相互作用。各种不同类型的蛋白质都会竞相结合到生物材料表面的粘附位点上。根据所附着蛋白质的不同构象，各种各样的信息可能传向附近细胞。本章后面会介绍通过改变生物材料表面来控制组织与生物材料之间相互作用的方法。

同种异体移植时，如果移植物的主要组织相容性复合物（Major Histocompatibility Complex，MHC）与受体的类型不同，那么就可能出现急性排斥反应。MHC 是一类细胞表面分子，它们为 T 细胞提供有关已被识别的异物的信息。在不匹配的两组 MHC 作用下，T 细胞会收到两套有关异物的指令，从而导致极其强烈的免疫反应。组织配型可以减小这种排斥作用，虽然患者通常仍然需要长期服药，用于抑制

免疫系统的某些活动。人体对组织工程中使用的生物材料制作的支架也会产生排斥反应。支架中的植入细胞会被识别为异物，并在巨噬细胞等免疫细胞的直接攻击下遭到破坏。或者在保护寄主的炎症反应过程中，植入支架外围会生成一层厚纤维膜囊，由于这种膜囊阻断了营养的供给，支架内的植入细胞会被饿死。因此，某些组织工程器件将植入细胞包在选择性通透膜材料内，以防止免疫系统的攻击，例如，用于释放胰岛素的海藻酸水凝胶颗粒就是如此。

金属植入体的腐蚀会释放出金属离子，从而可能造成某些人体的金属过敏反应。这种过敏反应会降低骨骼融合的速度，导致骨骼融合不充分，或者产生皮炎。一旦发生这两种情况，通常都必须取走植入体。这又一次说明生物材料并非惰性的。

但是，有时我们却需要有意设计生物材料来增强免疫系统的反应。例如，疫苗中一般都加入了一种称为佐剂的生物材料微粒，以提高并延长其免疫作用。疫苗佐剂可以具有自己的免疫特性，对免疫系统产生较强的局部刺激；也可以只是吸附弱毒免疫原的简单微粒，用于增加弱毒免疫原的效率，并增强巨噬细胞的吞噬作用；佐剂还可以是受控的缓释载体，能够延长局部区域的弱毒免疫原的作用时间，提高局部免疫反应的几率。选用疫苗佐剂时要权衡其副作用。美国食品与药物管理局（FDA）已批准的用于人体的材料只有磷酸铝、氢氧化铝和磷酸钙 3 种，它们都是矿物盐。科学文献中常将铝化物误认为明矾，其实明矾是硫酸铝钾。用磷酸铝或氢氧化铝沉淀抗原时，明矾是初始溶液。其他用于科学研究的生物材料佐剂还有乳化油、细菌脂多糖（LPS）以及它们的各种合成衍生物（即脂质体）。

6.5　利用仿生材料引导组织修复

6.5.1　表面化学修饰

细胞和组织与生物材料表面的相互作用是促进新生组织的沉积及其与周围胞外基质的健康融合的关键。如图 6.14 所示，一旦生物材料置于体内，材料表面上发生的许多分子水平上的化学反应都会影响细胞的响应。例如，在传统金属植入体表面，金属离子和氧离子会逐渐扩散出来，并与钙、磷、钠、氯等生物体离子相结合。同时，天然生物分子会在植入体表面不断吸附和脱离。正如 6.4 节所述，某些蛋白质在植入体表面的吸附会引发一连串事件，如血液凝结、炎症细胞聚集、细胞分化等。为了获得理想的响应，人体内原本每个细胞表面的化学组成和胞外基质都受到严格的调控。细胞与胞外基质之间的相互作用由细胞表面的受体（即整合素）与细胞粘附蛋白（即配体）的作用决定，其中细胞粘附蛋白质与胞外基质相结合。在植入器件应用中，生物材料取代了胞外基质，并通过同样的机制向作用于其表面的细胞发送信息（见图 6.14 ）。因此，生物材料表面在决定组织与材料之间的相互作用中具有非常重要的作用，这个概念对于生物材料科学家非常重要。

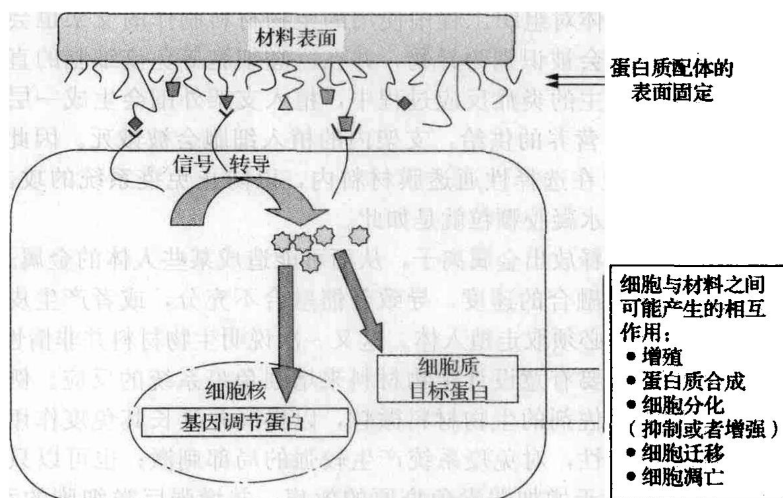


图 6.14　与细胞外基质相似，粘附了蛋白质的生物材料
可以通过配体与受体结合的机制向细胞发送信息

注：第一细胞信使经过多个途径的信号传导得到增强，导致各种细胞内信号分子的合成。这些信号分子通过对细胞核内基因调节蛋白质的作用以及对细胞质中其他目标蛋白质的作用，诱导各种基因的表达。理论上，生物材料科学家可以通过材料表面的适当处理来激发理想的细胞活性。

控制细胞与材料表面相互作用的一种仿生方法就是预先在植入体表面吸附一些蛋白质分子，模拟那些参与细胞粘附的分子的作用。现在大家已经知道，纤维连接蛋白和骨涎蛋白中存在的精氨酸-甘氨酸-天门冬氨酸（Arg-Gly-Asp，RGD）三肽序列可以调节细胞与材料表面的粘附性，因此，常在材料表面沉积包含 RGD 的多肽分子，以促进细胞的粘附。通常以蛋白聚糖形式存在的肝素和硫酸类肝素对于细胞的粘附也具有促进作用。另外，在植入体表面吸附生长因子等其他生物分子可以调控组织与材料之间的相互作用，从而增强细胞的活性。与只是单独添加了促细胞粘附分子的表面相比，这种添加了生长因子的表面可以促使更多细胞分化。

在生物材料表面粘附仿生多肽序列的化学方法有好几种。例如，可以通过所谓的有机硅化学法将蛋白质固定到材料表面上（见图 6.15a），这里不详细介绍其中的化学偶联反应和衍生过程。简而言之，这种方法是用硅烷偶联剂在生物材料表面与粘附蛋白质之间建立共价结合，由此形成有序的蛋白质粘附。用于建立偶联基团的固体表面修饰技术有很多种，如光化学修饰、化学衍生物、气体放电等离子体等。利用物理吸附原理还可以产生其他固定蛋白质的方法，如范德华力和静电吸附等方法（见图 6.15b）。静电吸附这种物理方法是

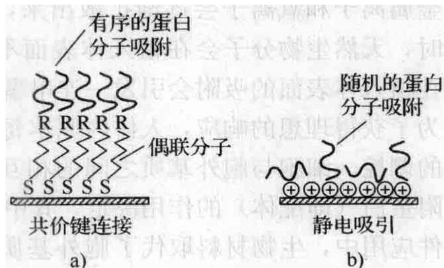


图 6.15　生物材料表面的两种修饰方法
a）共价键耦联技术
b）利用静电作用力的物理吸附方法

最简单的技术，它的特异性最差，且吸附的分子容易脱落。脂质基团和染料分子也可以用于蛋白质的表面固定。

表面修饰的一个关键指标是配体的密度，这种密度可以控制。例如，如果偏惰性的聚合物表面的蛋白质吸附性太弱，那么经过等离子体放电处理之后，材料表面就可以粘附较多的功能基团。等离子体处理后，材料表面具有较高的表面能，其反应活性增强，一般就可以增加生物组织的粘附。

利用表面修饰技术还可以产生抗蛋白粘附的材料表面，用于血管支架等与血液接触的移植体。例如，在材料表面贴上聚氧化乙烯就可以减少蛋白质的吸附。细胞在这种处理过的表面上的粘附明显减少。生物材料表面附上抗凝血剂也可以减少不想要的细胞粘附。已经证明，各种亲水材料可以减少血小板的粘附和血栓形成，另外，亲水材料还可以阻止骨骼愈合。由此可见，适用于某些情况的生物材料并不一定适用于其他应用情况。

透明质酸（Hyaluronan）是软骨细胞胞外基质中的一种生物大分子，它负责将蛋白聚糖聚合起来并栓在胶原基质上。研究结果表明，透明质酸可以引导间充质干细胞分化成为软骨细胞。实验时，透明质酸通过化学键粘附在培养皿上，然后加上未分化细胞。结果发现，分子量在 200000～400000 道尔顿这个特定范围内的透明质酸最容易启动软骨的形成。酶、抗体、抗原、脂质、细胞表面受体、核酸、DNA、抗生素和抗癌药物等生物大分子都可以固定在聚合物、陶瓷或者金属表面，或者固定在这些材料的内部。

此外，材料表面沉积钙离子或者使用含钙材料都可以增强骨细胞在材料表面的吸附力。例如，在表面覆有羟基磷灰石涂层的髋关节植入物上，纤维组织的形成会明显减少，而直接的骨骼连接则明显增加。带羟基磷灰石涂层的牙科植入物和脊椎融合笼也具有较好的骨细胞粘附特性。最近的研究还表明，羟基磷灰石陶瓷对于骨髓细胞的吸引具有选择性，这可能是由于其中蛋白质吸附的选择性引起的。

6.5.2　表面拓扑形貌

有证据表明生物材料表面的拓扑形貌代表了许多细胞反应的诱发信号，包括细胞粘附的调控、细胞形态、细胞凋亡以及基因调控等信号。因此，生物材料表面的修饰可以对组织的生长产生巨大影响。早期的全关节置换体具有光滑的表面，而现在的关节植入物表面却是粗糙、多孔的涂层，或者是开了凹槽的表面，这是为了引导骨细胞向植入体内生长。图 6.5 和图 6.6 所示的髋关节和膝关节植入体表面就是粗糙多孔的涂层。多孔涂层是通过将金属小球颗粒融合在移植体表面形成的，颗粒之间的空隙就构成了小孔。表面的小孔和沟槽可以诱导骨细胞迁移进来，或者沿着这些结构生长。据观察，沿着这种大于 5μm 的表面结构延伸和排列的细胞会诱发骨骼的沉积，从而使骨组织与植入体之间形成机械联锁，增加了两者的连接强度。这种受到引导的细胞活动被称为接触导向，接触导向机制也被用于增强其他类型细胞在组织工程支架中的迁移。除了引导细胞向小孔迁移之外，不同的表面拓扑形貌还可

以用于限制细胞的蔓延，迫使细胞的生长返回自身方向。例如，通过限制细胞在狭长的路径中生长，内皮细胞可以被诱导成为向上生长，从而形成三维的毛细管。

如图 6.16 所示，利用光刻技术可以在材料表面印制出各种蛋白质、分子或功能基团的微型图案。其方法简述如下：材料表面先涂上一层光刻胶（即光致抗蚀剂），再盖上掩膜，有选择地曝光某些区域之后，就可以获得特定的图案。当光照穿过掩膜的空隙，降解了暴露部分的光刻胶之后，这些部位就露出了原生物材料表面。于是，蛋白质等分子就可以粘附到这些暴露的区域上。然后，再去除剩余的光刻胶，就可以得到印制了蛋白质等分子图案的生物材料表面。相关细胞培养实验已经证明，特定的细胞容易粘附到经过化学修饰的区域。改变材料表面拓扑形貌的其他方法还有利用激光消融或者腐蚀剂的湿法蚀刻，这些技术可以使材料表面变得粗糙。据观察，为了对细胞生长和粘附能够产生一定的作用，材料表明的粗糙度必须在某个生物相关的尺度范围内，即 1～10μm。

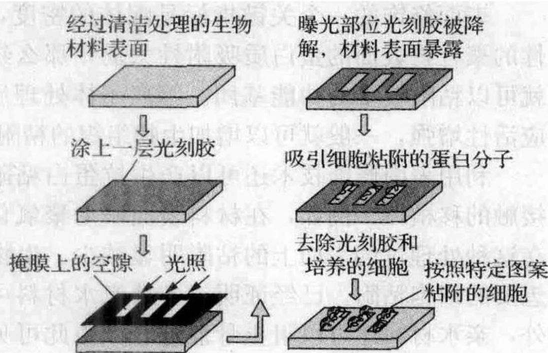


图 6.16 利用光刻技术可以在生物材料表面印制各种蛋白质或者功能基团的微型图案

具有光滑表面的材料，如能够抗蛋白和细胞粘附的热解碳等，是制作心脏瓣膜的理想材料。不过，如果用牛或猪的天然生物心脏瓣膜就更好，其理想的表面形貌和表面化学特性可以减少凝血和血栓的形成。但是，生物瓣膜一般会因钙化而失效。与光滑的热解碳表面相比，天然胶原表面比较容易钙化。

6.5.3 支架

修复较大的组织缺损时，最好用支架材料填补缺损空间。支架可以模拟胚胎细胞外基质的微环境。支架必须具备多种功能，必须能够提供生长因子附着的位点、细胞迁移和粘附的位置以及新生组织沉积的空间。天然组织是理想的支架选择，但是由于供体非常有限，因此人工合成的组织替换物的市场需求很大。虽然用螺钉固定的金属板或者用高分子聚合物板可以暂时将愈合组织固定在一起，但是这些材料都没有再生功能。要么用天然生物组织，要么人工模拟天然组织的细胞外基质及其细胞和生长因子，这样才能重建功能性的生命组织。支架的目标就是重建细胞微环境的要素，使细胞能够增殖、分化，并合成细胞外基质。目前市场上已有人工合成的再生性材料产品，包括携带或不携带刺激性生物分子、包含或不包含细胞的生物材料支架，它们一般都用陶瓷或者聚合物制作。生物材料支架最关键的性能是必须能够模拟细胞外基质的功能，能够支撑起空间，支持细胞，并且将细胞组合起来，构成组织。第 6.5.1 节讲述表面化学时已经介绍过模拟胞外基质中的粘性蛋白可以

促进细胞的粘附和分化。本节的重点是讲述胞外基质支架的结构和物理特性，它们对于促进细胞生成功能性再生组织具有非常关键的作用。

用于组织再生的生物材料支架中包含的孔隙的大小是一个很重要的参数。经过大量的试验和失败，现在人们已经确定了用于不同组织的各种生物材料的最佳孔径范围。根据现有的基本经验原则，细胞能够通过的孔径至少有 5～10μm 大小；较成功的骨骼支架一般在整个支架厚度上全都布满孔隙，大小约为 100～250μm；新生血管聚合物支架中的孔隙一般为 0.8～8μm，聚合物内的孔隙的孔径不能小于 0.02μm。通常，聚合物中孔隙的孔径可以比陶瓷或金属材料中的孔径小一些，这可能是由于聚合物在体内的降解和膨胀会引起孔隙的扩张。

支架孔隙的孔径大小决定了其多方面性能，如机械强度，对气体、液体和营养物质的渗透性，以及细胞向支架内部的生长等。多孔的交织结构是组织工程器械中营养物质扩散和组织细胞生长的基础。高孔隙率材料的降解速度要比固体实心材料快。将材料长时间浸泡在模拟体液中，一般可以测得支架材料的生物降解速率。

生产多孔材料的方法有好几种。有些方法比较适合生产用于药物输送的材料或构建低密度的硬质加固器械，但并不适合生产组织工程中使用的细胞支架。在材料中生成孔隙的一种方法是使用致孔剂。也就是，用溶剂将聚合物溶解，并与某种颗粒物质（即致孔剂）混合在一起，此时致孔剂在聚合物溶剂中是稳定的，不会被溶解，但此后可以溶解除去。然后，等到溶剂挥发以后，就留下含有致孔剂的聚合物薄膜，就像含有坚果的巧克力糖一样。再用水等其他溶剂将致孔剂溶解洗脱，聚合物中就形成了孔隙。这种方法称为溶剂浇铸一颗粒沥滤（solvent casting/particulate leaching）技术。与该技术相关的还有凝聚（coascervation）技术，这种方法先将聚合物溶解在溶剂中，再加入致孔剂，然后再添加不能溶解聚合物的溶剂。于是，聚合物就会沉淀出来，其中包埋了致孔剂颗粒。收集沉淀物之后压实，然后洗去致孔剂，就可以得到多孔材料。图 6.17 所示的骨晶体与聚羟基丁酸一羟基戊酸共聚酯（Poly

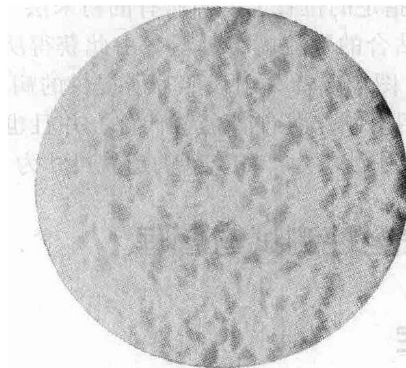


图 6.17　骨晶体与可吸收 PHBHV 制作的复合生物材料支架

注：生产这种支架时应用了凝聚技术，将聚合物从溶剂中沉淀出来，并使用了可溶性致孔剂形成骨骼修复所需的大孔隙。

Hydroxy Butyrate Hydroxy Valerate，PHBHV）的复合材料就是这么制作的。这些方法中都使用有机溶剂，不会在制作过程中带入药剂的污染。不过，致孔剂的洗脱大大增加了支架制备的时间。因此，人们还开发了气体发泡等其他方法。其中有一种气体发泡法的制备过程如下：将聚合物制作成固体圆片，在高压二氧化碳气体中放置 3 天，然后迅速取出圆片。当这些含有气体的圆片突然回到正常大气压下时，其中的气体就膨胀成气泡，有的气泡大小可达 100μm，并且材料中的孔隙率可以高达 93%。其他气体发泡法还有添加化学发泡剂，加热后即可产生气泡。但是，气体发泡技术形成的孔隙之间往往不相互联通，使得细胞种植和迁移都很困难。

相分离与乳化技术也可以用于制造高分子多孔支架，其方法简介如下：先用溶剂将聚合物溶解，再加入一种不相溶的液体（如水），混合成乳液。然后，将这种聚合物与水的混合液倒入模具，快速冷冻干燥，这种方法称为冻干。这样，原来被水占据的空间就变成了孔隙。孔隙率在 95%以上的高孔率支架就是用这种方法制成的。不过，这种方法产生的孔隙较小，只有 13～35μm，这是一个缺点。还有一种制作多孔材料的方法是纤维粘接法，也就是将纤维一层层排好或者编织起来，然后加热熔化后将它们粘连在一起，或者用溶剂粘接。这种方法的优点是孔的大小可控制，而且孔隙之间相互联通。缺点是孔道是规则的矩形，不像天然细胞外基质的结构。此外，加工过程中的高温处理和溶剂处理都不允许材料中结合生物活性分子。

生物材料支架还可以用实体自由成型制造（Solid Freeform Fabrication，SFF）技术制作，这种技术常用于汽车零部件的快速成型。其中，光固化立体成形法（Stereolithography）是利用一束聚焦激光的扫描，根据计算机辅助设计的指令，有选择地一层层固化聚合物溶液薄层中的某些区域。一层固化完成之后，固化处溶液的底部升高，在计算机控制下，激光束重新沿新的途径固化部件的下一层；如此重复，就可以制造出复杂的三维形状。同样，所谓的三维印刷术可以用于制作多孔支架，也就是一层叠一层地制造出复杂的形状。这种方法根据计算机指令在粉末池中通过“打印头”将“胶水”滴到指定的位置上。当所有的粉末层“打印”完毕之后，提起粘合在一起的部件，没有粘合的颗粒就会掉落，由此获得所需形状的三维部件。根据计算机辅助断层扫描 CT 图像或者 X 射线照片所提供的病人骨骼缺损部位的形状，上述制造方法已经成功地用于制作各种骨骼替代件，并且也用于制备各种聚合物支架和陶瓷支架，这些制造技术能够生产的支架的形貌尺寸为 10～1000μm。

6.6 生物材料的安全性测试与管理

6.6.1 产品特性的测定

生产安全产品的第一步是首先必须确保产品的纯度和特性。有关各种材料安全性的数据非常多，但是，由于各种加工方法可能引入添加剂，而且最终的灭菌消毒处理也可能改变生物材料的特性，因此，必须对最终产品的纯度和特性进行检验。

美国材料与试验协会（American Society for Testing and Materials，ASTM）已经颁布了很多标准，医疗器械生产厂商可以根据这些标准评价产品的纯度和特性，以及产品的安全性。ASTM 的网址是 www.astm.org。ASTM 对于用于医疗器械的金属有严格的规定，包括组成成分、微结构、晶粒的大小、杂质的大小、缺损的大小以及宏观孔隙率和微观孔隙率等都有相关的规格，以确保产品的安全。ASTM 有关陶瓷材料的标准规定了材料的化学成分、相位测定、晶粒大小以及可能降低材料抗疲劳性能的烧结辅料等杂质的含量。而且，美国 ASTM 正在编写新的标准，用以确保组织工程支架材料的质量和可靠性。除了规定支架材料化学特性的正确测试方法以外，ASTM 还制定了测试材料孔隙率和渗透性的标准方法。此外，ASTM 对于产品的溶解性、降解性和稳定性等测试也有相关标准。ASTM 的各项标准是在广泛收集意见的基础上撰写的，代表了制造商、用户等广大利益群体的意见和共识。

评估生物材料的组成和结构的方法有很多。不过，没有一种技术单独能够提供全部所需的信息，只有使用多种分析方法才能获得全面的信息。例如，晶体材料可以用 X 射线衍射技术来鉴定（一般分析样品都呈粉末状）。其原理是：X 射线束穿过一排排原子时会发生衍射，如果用检测器在材料表面不同角度测量反射射线的强度，那么，射线强度随测量角度变化的函数就体现了晶体结构的特征。将测量结果与现有的数据文档进行比较，就可以鉴别产品的晶体结构。

傅里叶变换红外光谱法是 X 射线衍射晶体结构测定的一种补充方法，它可以测量材料的化学基团数据，而且，被测材料不必是晶体，气体、液体或固体都可以。红外光谱法检测的是样品在红外线照射下所产生的分子振动。如果红外辐射的频率与材料中分子的基本振动模式相匹配，那么，红外辐射的能量就会被吸收，使得振动偶极子垂直于材料表面。红外辐射吸收系数随波长变化的测量曲线就构成了所谓的化学“指纹”，将这种指纹与现有的数据库中的各种材料指纹数据进行对比，就可以识别材料的类型。多数未知材料鉴别方法都是利用这种形式，也就是利用某种能量轰击材料，定量测定材料的作用结果，然后在数据库中搜寻类似的结果。例如，二次离子质谱法（Secondary Ion Mass Spectroscopy，SIMS）和 X 射线光电子能谱法（X-ray Photoelectron Spectroscopy，XPS）等技术都是利用这种基本测定模式，这两种技术常用于生物材料表面化学特性的测定。

扫描电子显微镜（Scanning Electron Microscopy，SEM）对于确定生物材料的二维表面拓扑形貌非常有用。其原理简介如下：用一束高能电子束扫描样品，使材料表面发射二次电子，这种二次电子的强度主要取决于材料表面的微观形貌；因此，通过记录二次电子所产生的电流，就可以重建测量表面的图像。SEM 的放大倍数可达 100000 倍，如果需要更高的放大倍数，还可以用原子力显微镜（Atomic Force Microscopy，AFM）。AFM 的原理是，利用固定在悬臂上的原子级的探针扫描物体表面，扫描时探针并不与物体发生实际接触。被测材料表面的原子与探针相互作用，互相之间会产生排斥力或者吸引力，这种作用力使得悬臂高度发生非常微小的变化，通过记录这种高度变化就可以构建材料表面形貌的图像。如果操作正确，AFM 的图

像可以显示出单个原子。

计算机辅助X射线断层扫描（即显微CT）和磁共振成像（NMR）技术可用于采集材料的三维图像，图6.18所示就是生物可降解多孔支架的显微CT图像。这两种三维成像技术需要在样品各个不同方向上进行扫描，然后将所有方向上采集的数据整合起来，通过数学算法重建三维图像。它们可以无损检测固体材料的孔隙率和体积分数，还可以直接测量固体的孔隙尺寸参数（如宽度、直径、厚度等）以及重复性结构的间距等。显微CT图像的对比度由X射线的衰减决定；因此，具有较高衰减系数的聚合物等支架材料可以产生对比度较高的图像，图像中支架与背景之间的自动分割计算就比较容易实现。目前市场上出售的台式显微CT产品的空间分辨率约为5μm。

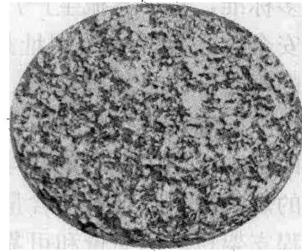


图6.18 图6.17所示支架的显微CT图像（由Connecticut大学医疗保健中心显微CT研究室的Douglas J. Adams博士馈赠）

6.6.2 产品安全性与生物相容性的测试和评估方法

本章前面已经反复强调人体内环境对生物材料的影响，然而，反过来，生物材料也会对人体产生影响；因此，在生物材料植入人体之前必须研究这两方面的作用，这就是所谓的生物相容性和安全性测试。进行这些测试之前，必须先应用6.6.1节已经介绍过的方法全面测定材料的结构和化学组成，以便确认材料的纯度和特性，并确保在合成、制造和消毒等各个处理环节中没有无意间引入其他物质。

生物相容性与安全性测试包括细胞和组织的离体试验、动物的在体试验以及最后的人体临床试验，美国ASTM和国际标准组织（International Standards Organization，ISO）以及FDA等联邦政府的监管机构都对此制定了许多指南和规则。ASTM的F-748标准和ISO的10993条例都给定了全面的安全性和生物相容性测试的详细方法，医疗植入器械的生产厂商必须遵循这些标准和方法。所需的测试根据产品的不同用途分类，表面器械、外部通信辅助设备、植入器械等产品的测试各不相同。根据与人体接触时间的长短不同，各类产品的测试还进一步分为短期接触、长期接触和永久性接触3类。基本测试包括细胞毒性、致敏性、刺激性（即皮内反应）、急性全身毒性、亚急性全身毒性、遗传毒性、植入试验、血液相容性、慢性毒性以及致癌性等。试验样品最好是完整的医疗器械，并且经过消毒等真正与用于人体时一样的处理过程。不过，由于试验条件的限制，有时不可能使用完整的医疗器械进行测试，因此，经常先将可滤取的成分或者植入物体的可降解成分进行离体试验（即细胞毒性试验），并且进行初步的在体试验（即致敏性和急性全身毒性、亚急性全身毒性试验）。要完成F-748和ISO 10993所规定的全部临床应用前试验一般需要2年时间，即使合格的、有经历的器械也是如此。

由于在体系统非常复杂，涉及很多变量，因此，离体测试结果与在体测试结果很难一致。一般，细胞培养实验比在体试验敏感；因此，离体试验显示毒性的材料并不一定不能用于人体。另外，细胞培养试验既可能出现假阴性也可能出现假阳性，因此，测定产品安全性和生物相容性时必须进行动物试验，同时为了初步评估产品的功能及其有效性，也需要进行动物试验。但是，不同种类的动物对于各种生物材料和药物的响应各不相同；例如，豚鼠对免疫过敏试验最敏感，而兔子则对在体热原试验最敏感。虽然动物试验的确是一种有用的筛选方法，可以防止有毒成分植入人体；但是，最终的生物相容性和安全性测试还是必须由人体临床试验完成。有时，在小鼠或者狗模型中被证明有效的产品在人体中并不一定有效，测试新药时尤其如此。药物的有效剂量有时在不同物种之间的差别相当大，在不同人群之间也会有差别。因此，在生物相容性、安全性和有效性测试中，要按照逐步增加剂量的方案进行，临床试验需要大量病人的参与。

6.6.3 管理规程

在美国，用于人体的新药、生物制品或者新的植入式医疗器械在实际使用和销售之前都必须获得食品与药物管理局（FDA）的批准。目前，FDA 共分为 6 大中心：器械与放射卫生中心（CDRH）、药品中心（CDER）、生物制品研究中心（CBER）、食品与化妆品中心、兽药中心以及毒理研究中心。FDA 在收到厂商的产品论证申请报告后，首先要确定该产品属于哪种类型，是药物、仪器还是生物制品；然后再决定由哪个中心主管其审查过程。FDA 每个中心有不同的处理程序和要求，必须完成这些程序并满足其要求，才能获得 FDA 的批准。有些产品是生物制品与器械的结合，或者是药品与器械的结合，这种产品必须满足两个以上 FDA 中心的相关要求才能取得许可证。例如，组织工程医疗产品就是多种形式组合的产物。这种组合型产品一般最难管理，进入市场之前的处理时间也最长。FDA 通常将不携带细胞和生长因子的单纯生物材料支架归为器械类产品，由 CDRH 负责管理。

由于 FDA 的认证程序太复杂，厂家一般都要聘用专家来处理新产品开发的 FDA 认证事宜。不过，开发人员也应该早就具备产品法规方面的知识，与良好的研究和开发工作一样，这也是医疗产品获得成功的必备条件。下面简单介绍 FDA 的器械与放射卫生中心（CDRH）的有关规程。

CDRH 根据医疗器械的分类来确定产品论证的要求和控制程序（参见网址 http：//www.fda.gov/cdrh/devadvice）。医疗器械产品分为三大类：Ⅰ类产品是那些与人体接触有限、风险最低的医疗器械；Ⅱ类产品需要某些特殊控制，为了确保其安全性，这类产品通常必须满足某些性能标准；Ⅲ类产品是不能明确那些基本控制或者特殊控制足以提供其安全性和有效性保证的产品。市场上还没有可比性产品的新型医疗器械自动列入第Ⅲ类产品。Ⅰ类产品有牙线、压舌板、外科医用手套以及妊娠试验等临床生化测试系统等；Ⅱ类产品有血压计袖套、氧气面具、牙科印模材料以及心电图机等；Ⅲ类产品则包括心脏瓣膜、自动血液加热装置以及硅胶乳房假

体等。产品分类是 FDA 认证过程中非常重要的一步，它决定了产品出售并用于人体之前必须接受哪些测试。

厂家至少在新产品或者具有重大改进的产品出售前 90 天向 FDA 提交上市前通告申请书。CDRH 收到的 99%以上申请书都可以通过 510（k）条例规定的上市前通告程序进行处理。该上市前通告的目的是请厂家向 FDA 证明新产品与已被批准销售的某种商品具有实质性的等同，也就是，确认所涉及物质的特性和纯度，然后根据质量体系法规简要地填写 ASTM F-748 和/或 ISO 10993 测试协议表。该法规还要求厂家提交产品的主工艺文件（Device Master Record，DMR）和设计历史文件（Design History File，DHF）。这两个文件中包括设计过程文档、制造细节、所有测试结果（包括校验测试）以及详细的设计原理和检验方法等。

如果新产品不符合 510（k）条例的规定范围，也就是目前已经合法上市的产品中不存在与该产品具有实质性等同的产品，那么，厂家就必须提交完整的上市前申请（Pre Market Application，PMA），它包括有关产品安全性和有效性的所有要求的临床前测试和临床试验数据。根据产品与人体和/或血液接触时间的长短，ISO 的 10993 条例中有一个决策树，它有助于确定产品需要进行哪些生物相容性和安全性测试。临床试验的管理很严格，以免受试者在不知情的情况下遭受很大的风险。产品在临床试验之前必须已经成功通过动物的离体和在体安全性试验，并具备详细的相关文件。人体临床试验一般分为 4 个阶段：安全性测试阶段、疗效测试阶段、双盲阳性产品对照的疗效评估阶段（这 3 个阶段大约需要 5 年时间才能完成）、最后是上市后的监督管理（也就是 FDA 批准后普通公众使用产品的情况分析）。开始临床试验之前，厂家必须申请研究性器械豁免认证（Investigational Device Exemption，IDE），提交的申请报告包括所有的制造程序和质量控制程序、临床研究计划，还有已经评审过该申报计划的审查委员会名单（参见第 2 章）。FDA 会在 30 天内批准或者否决 IDE 申请。一旦 PMA 申请程序完成之后，FDA 会在 180 天时间之内批准或否决该项申请。如果获得批准，产品就可以投入市场，并按照申请中说明的用途用于人体，产品标签上必须注明用途。如果产品要按照其他新的用途销售，那么在合法出售之前必须向 FDA 再递交另一份申请，并接受审查。如果读者希望进一步深入了解 FDA 各个中心的规程，可以查询网站 www. fda. gov 上有关 FDA 认证法规的指南文件。

6.7 生物材料应用举例

6.7.1 肌骨的修复

植入体生物材料部件的设计和选择应该基于所需恢复的受损或病变组织的生理功能。肌骨（musculoskeletal）组织的主要功能是提供支持器官的构架，并提供肢体运动所需的力量。骨骼、软骨、肌腱、韧带和肌肉都是肌骨组织的组成部分，它们

具有不同的功能和不同的生物特性，在植入体设计和材料选择时这些部分都必须分别考虑。其中，骨骼是惟一能够自发再生的组织，它一直处于不断重塑的状态，始终在优化自身的结构，以便更好地满足机体的需求。正是由于存在这种持续的细胞重塑活动，因此宇航员会在失重情况下迅速损失骨骼的质量。与骨骼不同，软骨不含细胞，其自我修复功能很有限；因此，软骨损伤往往是永久性且进行性的。软骨覆盖在关节中两块相对骨骼的表面，它的功能是减小关节运动时的摩擦。韧带不仅具有被动的关节约束功能，而且还为稳定关节的肌肉收缩提供电信号。

修复受损或病变的组织和器官时，最好使用自体移植或同种异体移植的材料作为供体，但是这些材料的来源有限，所以，经常退一步选择仿生的合成替代物。据证明，仿生磷酸钙材料（即羟基磷灰石）可以显著提高骨细胞的活性，并可以单独使用或者结合胶原蛋白等其他聚合物一起作为骨移植替代物。羟基磷灰石不仅可以促进骨细胞的粘附，而且还可以控制干细胞分化为成骨细胞。这种特性对于组织工程技术非常重要，组织工程的目的不仅要恢复功能，而且还要恢复实际的生物组织。在组织工程产品开发中，其材料选择比传统方法更复杂；一般地，其材料必须具有足够的机械强度，并且材料表面的化学成分要有利于细胞的粘附和增殖，还可能需要具备药物输送载体的功能，能够释放生长因子，而且一旦新生组织形成之后，植入的人工材料必须能够被吸收或产生生物降解。

目前，修复关节软骨的一种方法是将自体健康关节的一部分移植到清理过的缺损部位；另一种方法是用细胞填充缺损部位，这些细胞能够促进软骨和/或骨组织的生长。后一种方法最近已经获得 FDA 的准许，它使用的是病人自身的软骨细胞。治疗时，首先采集病人的软骨细胞，经过体外扩增之后，再将细胞注射到由外科手术在缺损部位构建的腔室内。不过，研究结果表明这种方法建立的软骨组织的降解速度比胚胎发育期形成的软骨要快，或许这种人工仿造的生物过程中缺少了某种关键的细胞信号分子。

组织工程软骨重建的研究正在不断发展。例如，有人将软骨细胞或干细胞种植在胶原蛋白、纤维蛋白和聚乳酸等软组织支架中，放在包含外源性生长因子的生物反应器中进行预培养，从而生产新生软骨组织。这种方法利用了生物材料支架，是对只用细胞的方法的一种改进。还有人开发了悬浮在透明质酸中的间充质干细胞混合物，拟用于直接注入到受损膝关节的半月板中。羟基磷灰石可以刺激干细胞分化成骨细胞，非常适合于各种骨应用，与此类似，胚胎细胞外基质中存在的透明质酸具有软骨诱导性和抗血管生成的潜能，表现出作为软骨组织生物材料支架的良好前景。

显然，仿生支架材料的功能不仅仅是为细胞提供惰性的结构支撑，如图 6.14 所示，如果选择正确，那么，这些支架还可以诱导细胞沿着特定的分化途径发育。此外，由于各种机械作用力在多数组织功能中都发挥着重要的作用，某些具备力学仿生特性的支架很可能具有其优越性。

长骨需要承载负荷，传统上，其修复一般都用金属材料。在目前科学家和工程

师们还没有最终完全了解如何设计各种全关节及其包含的所有组织成分的情况下，最好选用传统金属和聚合物（或陶瓷）制作的股骨、胫骨和髌骨等部件构成的假体（见图 6.5、图 6.6 和图 6.10）。根据目前科学技术的发展趋势，未来大多数肌骨损伤和缺损很可能都会在早期损伤较小的时候就得到治疗，并且，那些能够传送细胞和生长因子的可吸收性复合材料将成为骨骼、韧带和软骨的首选材料，而传统金属承重材料的作用会逐渐被削弱。它们将来可能只作为临时承载的支撑物，等到生物体内部新生组织再生成功之后，金属支撑件就会被撤除。

6.7.2 皮肤再生

生理学上人体最大的器官就是皮肤，它在维护体内平衡状态、提供免疫机能、协助感官反馈中发挥着关键的作用。显然，皮肤受伤后这些重要生理功能都会受到损害，使得机体脱水和受感染的危险性大大增加，烧伤就是一个极端的例子。以前，同种异体皮肤移植是很多皮肤损伤的主要治疗方法。由于皮肤移植会进一步造成损伤，人们一直在努力研究和开发皮肤的替代品。生物材料的应用提高了皮肤替代物的性能，使其可以真正替代原生皮肤的功能。

如图 6.19 所示，宏观上，皮肤是由真皮和表皮构成的双层器官，而表皮又有自己的二级结构。真皮由于含有胶原蛋白，主要负责形成皮肤的结构要素。此外，真皮细胞，特别是成纤维细胞，会产生表皮细胞正常增殖和分化所需的化学因子。真皮和表皮之间存在一层明显的基底膜。基底膜是一种细胞外蛋白质基质，它是表皮的基底细胞粘附与增殖的基础。实际上，评估人工皮肤的常用方法就是检验其具备的基底膜的性能，因为它反映了真皮与表皮之间的结合。基底膜外粘附的第一层细胞就是基底细胞，这些细胞会不断增殖和分化，从而构成整个表皮层。按照细胞的不同分化程度，从内到外，表皮层排列的细胞层有基底细胞、有棘细胞、颗粒细胞

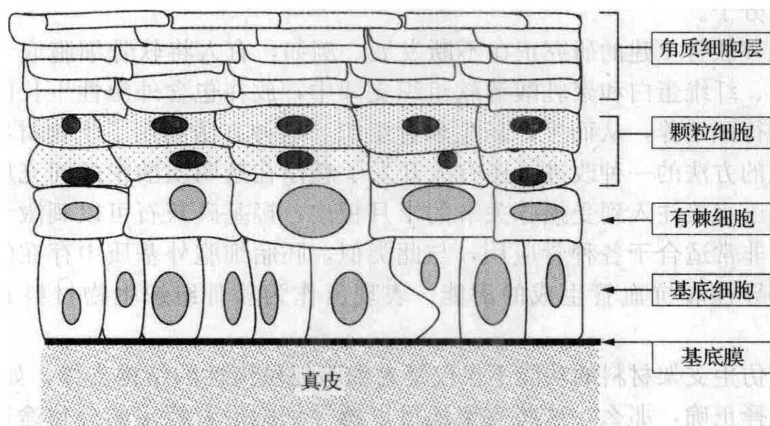


图 6.19 皮肤是由真皮和表皮构成的双层器官

注：真皮的主要细胞是成纤维细胞，而表皮的主要细胞组成是处于不同分化阶段的表皮细胞。此图由 Venkatasu Seetharaman 绘制。

和角质细胞层。其中，角质层由高度交联的角化细胞构成，是皮肤最外层屏障，它提供防脱水保护。

根据仿生学理论，有效的人工皮肤应该模仿天然皮肤的结构。而且，设计良好的皮肤替代品要具备物理上的完整性，可以抵御各种环境因素的作用，同时能够支持细胞的生长。图 6.20 所示是人工皮肤的示意图，其生物材料除了要为细胞的生长提供合适的基质之外，还要具有很好的机械强度。因此，作为一种交联聚合物，胶原蛋白成为一般人工皮肤的主要真皮替代品。胶原蛋白互相交联起来后就形成了一种力学性能良好的基质材料，在其上面可以种植表皮细胞，甚至还可以在其内部包埋成纤维细胞，给表皮细胞的正常分化和增殖提供化学因子。

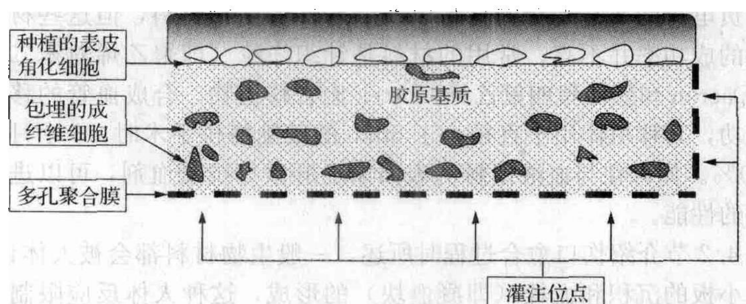


图 6.20　与天然皮肤具有相似结构的人工皮肤示意图（此图由 Venkatasu Seetharaman 绘制）

牛源 I 型胶原蛋白常用于人工皮肤，经过 pH 值和湿度的改变，它可以达到最优生理状态。将制备好的胶原蛋白溶液注入样本盘，使其聚合成为凝胶状物体，聚合过程的收缩会使胶原纤维相互交联在一起。通过调节胶原的制备方式和聚合时间的长短，可以控制凝胶的坚固程度。值得注意的是胶原凝胶（即胶原海绵）对于大多数分子都具有很好的渗透性，允许主要化学因子的扩散。因此，将成纤维细胞包埋在这种胶原纤维的交联结构内，就可以给表皮提供各种化学因子。为了构建具有完整双层结构的人工皮肤，还要在这种胶原蛋白构成的真皮层上种植表皮角化细胞，并促使其分化成适当的上皮细胞排列形态。还有一种人工皮肤用硅胶薄层作为功能性上皮屏障，而将真皮层直接敷于伤口上。后续处理再用表皮基质取代硅胶，进而完成整个皮肤的植入过程。上述胶原蛋白和硅胶仅仅是用于重建正常皮肤功能的两种生物材料的例子。

6.7.3　心血管器械

与循环系统的血液直接接触的生物材料器械，如心脏瓣膜、血管置换物或血管支架等，主要必须抗血小板凝聚和血栓沉积，并且具有一定的机械强度和耐久性，具有良好的生物相容性且无毒。这些要求与被替换组织的主要功能相关。例如，动脉由 3 层不同的功能层构成：内膜、中膜和外膜。内膜是血管的内层，它由内皮细胞组成，具有抗血栓表面，可以防止血液与中膜的凝血酶相接触。血管内膜的细胞

可以产生很多种生物大分子，包括各种生长因子、血管活性因子和粘附因子等。血管中膜（即实质组织）由多层有序排列的平滑肌细胞组成，是血管的中间肌肉层，它使血管具有所需的强度并保持血管的弹性。血管外膜就像一层硬护套，它保护中膜层的平滑肌免受过度的拉伸或者压缩负荷的作用。

与大多数组织一样，自体移植是血管组织更换的首选生物材料。例如，病人自身的静脉血管可以取出一段，用于替换堵塞的动脉。不过，静脉移植每年的失败率高达20%。由于大约30%的病人无法进行自体静脉移植，因此，人们开发了许多人工合成的移植材料。天然血管完好的内膜不会引起凝血，通过模仿内膜内皮细胞的某些特性，科学家和工程师们制造了具有较好血液相容性的材料。例如，有人设计了表面带有负电荷的非常光滑的材料以及亲水性的生物材料，但这些材料在血液接触性应用中的成功率并不高。常用的材料是针织涤纶（即聚乙烯对苯二甲酸酯）产品 Knitted Dacron 和膨体聚四氟乙烯 Gortex 血管移植物。合成血管的移植已经取得了一定的成功，其移植体用于直径小于6mm的动脉搭桥手术时，6个月形成血栓的概率约为40%。如果在与血液接触的表面涂上肝素等抗凝血剂，可以进一步改善这类血管产品的性能。

正如6.4.2节介绍伤口愈合过程时所述，一般生物材料都会被人体识别为异物，从而导致血小板的沉积和血栓（即凝血块）的形成，这种人体反应限制了生物材料在血液接触性应用中的发展。因而，组织工程学关于植入和构建多层活细胞结构的方法在心血管应用中展现了广阔的应用前景。根据组织工程的研究结果，人们最近才开始认识到，选择合适的细胞培养条件和支架材料可以促进平滑肌细胞排列起来，并紧固内皮细胞，从而生成血管。不过，要生产具有一定功效的组织工程动脉血管产品，还需要进行大量的研究工作。例如，虽然这种组织工程人工动脉可以抗血栓形成，但是，其高度细胞化的血管所具备的机械强度却不足以承受心脏搏动时产生的冲击力。

病变的人体心脏瓣膜目前可以用合成材料制作的机械假体或者天然生物组织制作的生物假体来替换。图6.21所示的3种人工瓣膜中有2种是机械假体，一种是生物假体。其中机械心脏瓣膜有4个基本组成部分：碟形或者球形阀体；阀门关闭时抵住阀体的基环；限制阀体活动范围的瓣笼；可以将人工瓣膜与心脏缝合在一起的缝合环。每次心脏搏动时阀体在基环与瓣笼顶部之间来回弹跳。阀体的活动越灵活，瓣膜的开关就越有效。因此，阀体的重量和阀体材料的耐磨性非常重要。早期的心脏瓣膜用较轻的塑料或者空心金属球作为阀体，现在已经证明硅胶制作的球形阀体很有效。不过，代替二尖瓣的碟形瓣膜要求材料具有一定的硬度，而硅胶制作的碟形瓣太软。聚甲醛（Poly Oxy Methylene，POM）要比硅胶坚硬，因此以前曾经被用于碟形阀体的制作。但是，其磨损却是一个问题。碟形阀体会在瓣笼中自由旋转，使碟阀的边缘受到均匀磨损。此外，碟阀在瓣笼中上下活动时，其边缘会产生磨损痕迹，且上下活动会阻止碟阀的旋转，使磨损痕迹加深，最终导致阀体失效。热解碳具有很好的耐疲劳强度和耐磨性，因此，目前它是制作阀体的主要材料

之一。



图 6.21　用于替代人体病变或畸形心脏瓣膜的 3 种人工瓣膜
注：生物组织瓣膜通常用猪的心脏叶瓣制作，笼碟瓣和笼球瓣则完全用人工材料制作。

为了保持外形，瓣笼和基环需要一定的强度和硬度。并且，在装入阀体时它们还必须能够被撑开。因此，这些部件一般都用金属制作，要么通过机械加工而成，要么将零件焊接而成。早期的人工瓣膜用钴合金制作瓣笼和基环，因为钴合金的强度和耐腐蚀性都较好。最近，已经开始用钛合金，原因之一是由于钛合金比钴合金要轻。另外，长期使用中也有人反应钴合金瓣笼会引起过敏反应。

人工瓣膜的阀体和瓣笼等都直接与血液接触。与异物材料的接触可能导致血液的凝固，引起血小板聚集和凝血。而且，血液流过瓣膜时会产生漩涡或者停滞，这样，凝血因子就会积累起来，并导致血栓的形成。因此，在选择材料和设计瓣膜结构时都必须考虑血液接触问题和血流问题。至今这些问题还没有完全解决，安装了机械心脏瓣膜的病人都需要服用药物来减小血凝块形成的危险。

生物心脏瓣膜的抗血栓性能明显优于合成材料制作的瓣膜，但它们不耐用。生物瓣膜一般都取自 7～12 个月大的猪心脏，然后保存在戊二醛固定剂中。这种固定剂会缓慢地使生物瓣膜产生不利的钙化，并最终导致失效。目前人们正在开发利用组织工程技术合成的活体心脏瓣膜，但其耐用性还是一个大问题。此外，心脏肌肉至今还从来没有移植过；不过，干细胞有可能在体原位重建受损组织。最近有实验结果表明，直接注入心室壁的间充质干细胞可以与心脏肌肉相结合，并诱导受损心脏肌肉的再生。谁知道这对子孙后代的预期寿命意味着什么！

6.7.4　药物输送

生物材料作为药品和生物大分子的输运载体发挥着重要的作用。制药界长期以来一直使用滑石粉和碳酸钙等粉状材料制作药丸和药片。药物输送（又称药物缓释）的研究目标是制备可以在人体内持续产生药物活性的给药剂型，从而提高药物的疗效。控释制剂就可以实现这个目标，它利用了药物与生物材料融合的各种技术。例如，将基本纤维细胞生长因子键合在肝素上进行输运时，其半衰期测量结果表明，药物在血液循环中保留的时间增加了 3 倍。另外，聚乙二醇（PolyEthylene Glycol，PEG）融合技术已经成为保持体内蛋白质稳定性的一种成熟办法。

静脉注射的脂质体药物制剂也能够使血液循环系统药物半衰期延长。通过磷脂在水环境内形成的疏水性和亲水性小室可以制备脂质体。单层脂质体由包围着水质核心的脂质双分子层构成。水溶性药物可以包在这种脂质体的核心中，而脂溶性药

物则可以溶解在脂质双分子层中。与普通药物剂型相比，脂质体剂型可以使血浆中药物的浓度升高 3～10 倍。

高分子聚合物也被广泛用于各种药物输送系统。例如，非降解的疏水聚合物（即硅胶）用得最多，它用于制作包裹药物存储池的半透膜。如图 6.22 所示，有一种制剂由药物与可吸收聚合物的混合物制成，当聚合物逐渐降解时，药物也就被缓慢释放出来。增加混合物中蛋白质等药物的含量就可以提高药物的释放速率。根据经验，混合物中聚合物的平均分子量越大，蛋白质等药物释放的速率就越慢。另一种控释制剂的制备方法是用聚合物包裹药物溶液，形成聚合物微球。如果聚合物遭受腐蚀而破裂，那么这种微球中包含的药物就会全部一起释放；如果聚合物缓慢地经受表面腐蚀，那么其中的药物就会缓慢释放。此外，利用细胞特异性配体还可以使药物只作用于特定的靶细胞。经过配体修饰的药物微粒不会在人体内均匀扩散，而是通过受体与配体的相互作用直接运输到某种细胞部位。例如，携带了叶酸配体的脂质体会被癌细胞优先吸收，因为与其他种类的细胞相比，癌细胞表达的叶酸受体比较多。

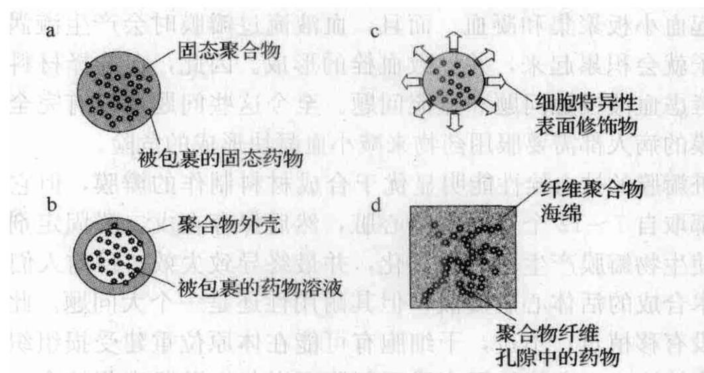


图 6.22 聚合物可以给药物和生物分子的控释提供各种基质

注：药剂可以装载在微粒、微胶囊、多孔聚合物中，或者与聚合物单链连接在一起。利用细胞特异性配体还可以设定药物的靶标。

将药物直接注射到目标组织中也是一种产生局部高药物浓度的方法。如果将药物与生物材料载体相融合，还可以实现药物在目标部位的控释。尤其当药物具有毒性副作用时，这种方法特别有效，因此此时只是将目标组织暴露在高浓度药物下，可以使副作用减到最小。如图 6.23 所示，这种局部药物输送方法可以降低血液药物浓度的峰值，并可以在较长时间内保持一定的药物活性水平，从而获得较好的疗效。除了局部直接注射以外，人们还开发了另一种携载化疗药物的磁性颗粒方法（如磁性脂质体）。静脉注射这种药物之后，在外加磁场的引导下，药物颗粒会聚集到癌症部位，而不会均匀扩散到全身。

有些药物不需要溶解就可以产生生物活性，例如，固定在材料表面的神经生长因子就具有活性。固定时引起的蛋白质构象变化有时甚至还会使表面固定的蛋白质

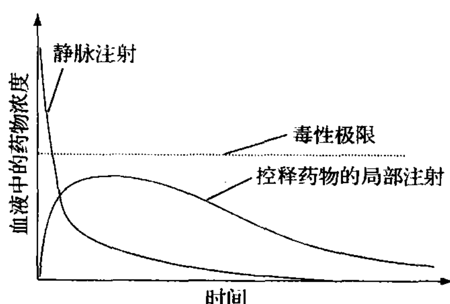


图 6.23　局部药物注射的主要优点是可以降低血液循环系统中的药物浓度，使浓度低于毒性水平，从而减少副作用；而注射部位或者植入部位的高浓度药物则可以增强药物的疗效，这种技术特别适用于化疗药物和激素治疗药物

获得更高的活性，这是由于固定之后蛋白质的活性部位暴露更多引起的。但是，相反地，固定后也可能使药物的所有生物活性都丧失。因此，在药物固定或融合到生物材料表面之后，必须进行各种测试，以便确定药物的活性。

胶原蛋白常用于骨骼生长因子的输送。这些蛋白质生长因子一般不与胶原蛋白表面形成共价连接，因此其释放速率较快。羟基磷灰石材料也用于运输骨骼生长因子，因为蛋白质等药物通常与羟基磷灰石表面可以形成很强的连接，使得其药物释放速率比胶原蛋白药物载体慢得多。根据不同的药物及其作用模式，释放时间长可能是优点，也可能缺点。

另外，目前正在开发的药物输送芯片与上述控释药物输运载体不同，它是一种可以通过外部信号打开储药微室的可编程载体。最巧妙的药物输送载体其实就是细胞。与独自发生作用的被动式药物输送装置不同，细胞会根据体内环境的信号产生细胞因子、生长因子以及细胞外基质物质。研究人员现在正在试图利用这种机制开发植入式的异种胰岛细胞封装体。胰岛细胞可以根据血液循环系统中的胰岛素浓度来产生胰岛素。异种细胞必须封装在藻酸盐等生物材料内移植，才能避免毒害移植细胞的机体免疫反应。目前该研究项目的重点是要改变藻酸盐的各种特性，使其具有足够的通透性可以维持其中包含的胰岛细胞的存活，还要能够保护细胞免受免疫作用的侵害，同时要允许胰岛素从植入体中扩散出去。看起来这对于材料的要求相当苛刻，医用材料的选择总是要面临各种特性之间的权衡。

习　　题

1. 请列出本章所述的 5 种基本生物材料，并给出简要说明。

2. 制作下列植入器械时你分别会选用哪种生物材料，为什么？请说明需要考虑的重要性质，为什么？同一种器械可以选用多种材料。

(a) 皮肤替代物；

(b) 神经再生的导管；

(c) 髋关节置换体的主干；

(d) 牙套；

(e) 导尿管；

(f) 组织工程骨。

3. 请列出制作以下5种器械的常用生物材料的化学名称：(a) 缝合线；(b) 心脏瓣膜；(c) 牙槽骨内牙根植入体；(d) 隐形眼镜；(e) 髋关节假体。

4. 在选择整形外科植入体的生物材料时应该考虑材料的哪些特性？

5. 如果骨组织中的碳酸磷灰石晶体的排列没有间隔，是连续的长纤维状，就像复合纤维那样，那么骨骼的力学特性会变成怎样？设骨骼矿物质的弹性模量是114～130GPa，皮质骨（即正常骨）的弹性模量是19～20GPa。

6. 请分别说出天然生物材料制作的医疗器械的三大优点和三大缺点。最常用的天然生物材料有哪些？请列出3种医学应用的例子。

7. 植入部位植入体与生物组织之间的反应情况是植入器械是否成功和稳定的重要反映。请列出4种可能发生的反应情况，首先说明有毒材料的情况。

8. 缝合的目的是什么？选择缝合线材料时需要考虑哪些重要特性？请列出4种常用的缝合线材料。

9. 请定义什么是钙化，哪类医学应用最可能由于生物材料的钙化而导致失效？

10. 什么是生物活性材料？什么是仿生材料？

11. 过去，如果植入材料在体内被纤维组织包裹，且没有引发进一步的人体反应，那么我们就认为这种植入材料具备生物相容性。为什么现在生物相容性的定义改变了呢？

12. 在植入体内后，为了使生物材料表面能够吸引细胞，需要在材料表面粘附某种蛋白质，特别是生长因子，为什么在这种情况下蛋白质的结构很重要？提示，请考虑生长因子是如何吸引细胞的，这些蛋白质又是如何能够固定在材料表面的，这样就可以说明其中的必要性。

13. 生物材料植入体附近会存在多种不同的细胞。请问植入体附近骨骼上会沉积哪些细胞？清除生物材料碎屑的又会是哪些细胞？

14. 请简述生物材料植入之后正常伤口愈合反应的4个阶段。

15. 假设你最近设计了一种新型生物植入材料，并进行了在体植入实验研究。当取出材料试样之后，发现材料周围包裹了一层纤维膜囊。此实验结果说明这种材料具有怎样的生物相容性，特别是其炎症反应如何？

16. 植入材料周围出现纤维组织层是不是对于所有应用都是好的迹象，为什么？哪些应用中纤维组织层是不希望产生的？

17. 请简述两种制作多孔生物材料的方法。

18. 什么仿生聚合物特别适合软骨细胞的分化，为什么？

19. 什么生物材料可以引导直接的骨结合，而不会产生纤维层的干扰？

20. 请简述一种方法，它能够将药物载体微粒（或纳米颗粒）修饰成以某种特定细胞作为其靶标。

21. 请根据 FDA 的规定（网址为 www. fda. gov），将下列生物医学器械分成Ⅰ类、Ⅱ类或Ⅲ类产品：（a）人工晶状体；（b）心脏瓣膜；（c）预制的齿冠；（d）氧气面罩；（e）听诊器；（f）牙科汞齐材料。

推荐阅读资料

Annual Book of ASTM Standards, Vol 13.01. (2004). Medical and Surgical Materials and Devices. American Society of Testing and Materials (ASTM) International, West Conshohocken, Pennsylvania.

Atala, A., Mooney, D., Vacanti, J.P. and Langer R. (1997). *Synthetic Biodegradable Polymer Scaffolds*. Birkhauser, Boston.

Dee, K.C., Puleo, D.A. and Bizios, R. (2002). *An Introduction to Tissue Biomaterial Interactions*. Wiley-Liss, Hoboken, New Jersey.

DeFrances, C.J. and Hall, M.J. (2004). *2002 National Hospital Discharge Survey: Advance Data from Vital and Health Statistics*, No. 342. National Center for Health Statistics, Hyattsville, Maryland.

Dillow, A.K. and Lowman, A.M. (2002). *Biomimetics Materials and Design*. Marcel Dekker, New York.

Helmus, M.N. (2003). *Biomaterials in the Design and Reliability of Medical Devices*. Landes Bioscience, Texas and Kluwer Academic, New York.

Hench, L.L. and Wilson, J. (1993). *An Introduction to Bioceramics*. (Advanced Series in Bioceramics, Vol.1.) World Scientific, River Edge, New Jersey.

Guilak, F., Butler, D.L., Goldstein, S.A. and Mooney D.J. (2003). *Functional Tissue Engineering*. Springer, New York.

Lanza, R.P., Langer, R. and Vacanti, J.P. (2000). *Principles of Tissue Engineering*, 2nd ed. Academic Press, San Diego, California.

Mann S. (1996). *Biomimetic Materials Chemistry*. VCH Publishers, New York.

Palsson, B.O. and Bhatia, S.N. (2004). *Tissue Engineering*. Pearson Prentice Hall, Upper Saddle River, New Jersey.

Park, J.B. and Bronzino, J.D. (2003). *Biomaterials Principles and Applications*. CRC, Boca Raton, Florida.

Saltzman, W.M. (2001). *Drug Delivery: Engineering Principles for Drug Therapy*. Oxford Univ. Press, New York.

Silver, F.H. and Christiansen, D.L. (1999). *Biomaterials Science and Biocompatibility*. Springer, New York.

Ratner, B.D., Hoffman, A.S., Schoen, F.J. and Lemons, J.E. (2004). *Biomaterials Science: An introduction to Materials in Medicine*. 2nd edition, Academic Press. San Diego, California.

Rho, J-Y., Kuhn-Spearing, L. and Zioupos, P. (1998). Mechanical properties and the hierarchical structure of bone. *Med. Engineering & Phys.* 20, 92–102.

Williams, D.F. Ed. (1987). Definitions in biomaterials. In *Progress in Biomedical Engineering*. 4, 67. Edited by D.F. Williams. Elsevier, Amsterdam.

第 7 章　组织工程学

Randall McClelland 博士
Robert Dennis 博士
Lola M. Reid 博士
Bernard Palsson 博士
Jeffery M. Macdonald 博士

本章目录

本章主要内容：

- 细胞治疗领域的发展；
- 3 类主要体外生物反应器；
- 组织功能的细胞动力学机制；
- 干细胞在组织功能中发挥的重要作用；
- 细胞发育过程的定量描述；
- 3 种不同结构中的物质传输分析；
- 表征组织环境的特征参数以及体外模拟组织环境的方法；
- 按比例放大的基本原理；
- 功能性组织工程的定义以及“组学”对于该新兴研究领域的促进作用；
- 细胞疗法和人工生物器官在临床应用中遇到的各种问题。

7.1　组织工程的定义

组织工程是生物医学工程的一个学科分支，它是生物学与工程学相结合的产物，其目标是体外（即离体）构建人体组织或者细胞产品，并且利用这些产品更好地进行体内（即在体）生物组织的修复。这门学科涉及许多生物学领域，包括细胞生物学、分子生物学、生理学、内分泌学等，以及细胞谱系的干细胞增殖与分化、胞外基质化学等；同时它还涉及许多工程领域，包括生化工程、机械工程、高分子科学、生物反应器的设计与应用、气体和液体代谢物的传输以及生物材料等。要将组织工程的研究成果转化成临床应用，还需要许多其他学科研究人员的参与，这样，才能使临床医生乐意接受并使用各种新型的工程化组织材料。这些学科的交叉结合产生了再生医学领域，该领域目前主要有两个临床战略目标：一是实现修复受损组织的细胞治疗，包括细胞或细胞悬液的注射和移植，有时植入物中还需结合支架材料；

二是离体构建用于移植的生物组织，或者制造体外器官用于辅助体内的病变器官。软骨、骨骼、皮肤、神经和肝脏等组织的细胞治疗和体外器官的临床试验现在已经起步，并且在近几年会迅速发展到许多其他组织的修复应用。不过，这种新型治疗模式是否能够取得成功仍有待于科学和经济两方面的评估。

7.1.1 组织工程师所面临的挑战

图 7.1 显示了组织工程师在细胞治疗或者植入体和人工器官构建时所面临的尺度上的基本问题。人造结构中的组织单元必须做得很小，能够仿真人体中天然毛细血管床的营养运输机制。这种体内毛细血管床通常由厚约 100μm 的组织薄片构成，其中包含着脉管系统；因此，为了给培养的组织提供适当浓度的营养成分，体外制作的人工替代品的微环境也必须符合这些条件。如果体外人造结构能够达到毛细血管床的这种尺度要求，那么，就可以进一步设计较大的工程化替代品给更大的细胞群体提供营养。图 7.1 所示的生物反应器就是具有合适的径向尺度，并且又能够在纵向扩展细胞空间的一种装置，其纵向长度约为 10cm。此外，无论是放大还是缩小培养规模，细胞培养的自动化装置中都必须保持正确的营养比例和生物物质相互作用的比例；因此，根据不同的需要，每一种特定的应用都具有独特的比例参数。

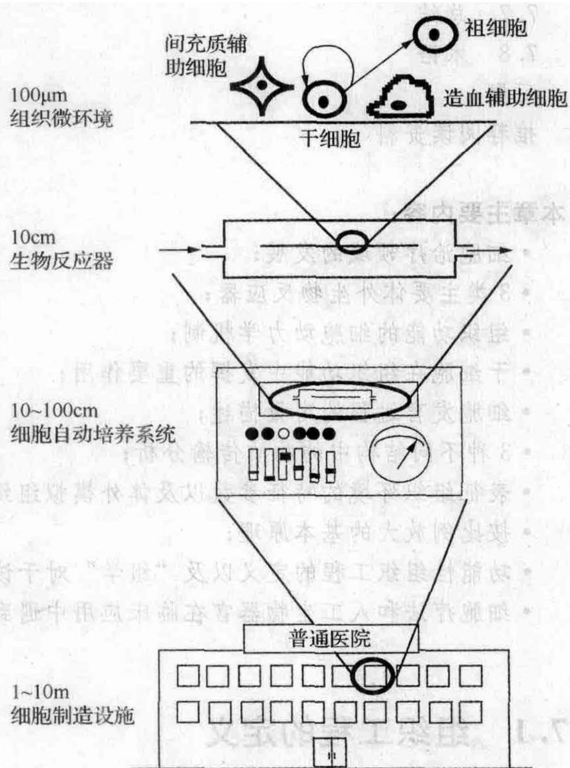


图 7.1 组织工程和细胞治疗中的 4 种主要尺度

下面将举例说明如下组织工程的研究方向：

1）为建立组织功能重构物理微环境和生物微环境，其中，物理微环境是指质量的传输，而生物微环境是指可溶性和不溶性信号分子。

2）构建具有临床意义的细胞微环境，解决比例放大所引起的各种问题。

3）在临床应用所需的比例尺度下实现系统的自动控制。

4）组织工程器械临床应用的实现，包括细胞治疗所需的细胞处理和保存方法等。

本章后面将重点介绍前两项研究，7.4 节也将简单介绍后两项研究所面临的问

题。为了使读者对前两项研究预先有个大概的了解，图 7.2 从人体内单个细胞的角度描述了生物组织的微环境。

图 7.2 的中间是一个细胞，然后由内向外表示了不同尺度下影响人体每个细胞的环境因素。这些环境因素主要可分为生物学因素（即化学因素）和物理学因素，其中，化学因素包括细胞外基质、激素、趋化因子和细胞因子、营养物质以及各种气体分子；物理因素则表现为几何结构、呼吸动力学过程以及新陈代谢产物的排泄过程。本章下面将首先详细讲述组织工程学中需要考虑的生物学因素，然后介绍物理学因素，最后，再综合细胞微环境的生物学和物理学因素，展望未来组织工程产品的开发研究与临床应用。

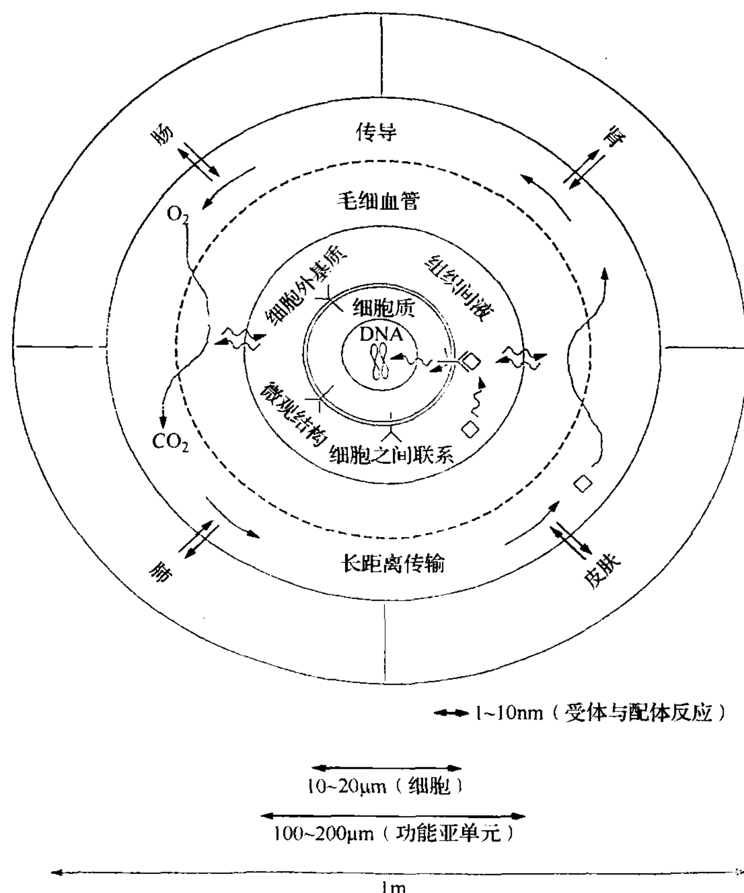


图 7.2 细胞及其与人体其他各部分的通信（根据 Lightfoot（1974）修订）

7.1.2 细胞治疗、移植物和体外人工器官

细胞治疗的发展最初源于细胞生物学和分子生物学的发展，但是，为了将这些领域的发展成果应用于临床治疗，必须开发一种技术，能够调控细胞培养并利用细

胞的代谢产物，因此，细胞治疗的实现需要生物工程、生化工程和生物材料学等许多学科领域的合作。目前细胞治疗所遇到的难题是需要设计组织或组织样的环境条件，以便能够培养足够多的细胞用于临床和产品生产。历史上，仅靠青霉素的发现并不足以影响人类的医疗保健，大规模生产技术的开发才是促进其临床应用发展的关键因素。抗生素大规模生产的实现无可置疑地说明了工程技术对于人类医疗保健发展的重大贡献。同理，开发人体细胞分离、扩增和低温贮藏的工业化方法也将使细胞治疗的应用常规化，满足临床需求。在这方面，组织工程学的研究重点就是各种人体组织的操作与设计。

本章后面将介绍细胞生物学的几个基本问题，包括细胞的分化、增殖、迁移和死亡（坏死或凋亡）等主要细胞生理过程；还将介绍干细胞和成熟细胞谱系的一些生物学基本知识，以及定向干细胞在器官功能的形成和修复中的作用。生物组织由多种细胞构成，各种细胞之间存在动态的相互作用；所以，只有将这些不同种类的细胞共同培养在一起，或者将某种特定细胞与来自其他种类细胞的信号因子结合在一起，才能表现出组织的特异性功能。这些信号因子包括细胞外基质中的不可溶因子、细胞与细胞之间直接联系的信号因子以及自分泌、旁分泌和内分泌作用产生的可溶性信号因子等。

这里，必须解决很多生物工程难题。例如，用于细胞治疗的注射针头的设计和注射程序的制定，其中，针头的设计必须尽可能减小细胞膜受到的剪应力。还必须通过分析营养物质的传输模式来计算维持组织存活所需的细胞集群的大小。此外，必须选择移植技术和种植位点，使植入的细胞能够在平衡的内环境中繁殖并发挥其作用。还要尽可能避免栓塞等不利事件的发生。在制造生物反应器等器具时，必须考虑细胞生长所需生物材料的功能、选材、制作方法和处理方法，以及器具的结构等。流体力学和物质传输在正常生物组织功能中发挥着重要的作用，因此，在体外细胞植入体的设计中也必须考虑这些方面的关键问题。另外，新陈代谢、细胞之间通信等细胞生理过程的系统性分析对于复制正常的人工器官和制定人工器官的技术指标也至关重要。性能良好的体外生物组织培养系统必须具备生物参量和物理化学参量的各种动态平衡状态，从而能够实现生物组织的功能。从数学上分析这些平衡状态时，可以采用无量纲的计算公式来描述各个时间常数的特征比值，也就是，建立新的无量纲参数用于表示“物理时间”与“生物时间”的比值等参量。

细胞治疗和移植体的临床应用的实现需要解决几大难题，其中包括新鲜组织的获取、细胞处理和分离、安全性试验、细胞活化或分化、培养基的制备、产品保存与稳定性以及产品的质量保证与质量控制等问题。本章将简单介绍这些难题。无论是现在还是未来，细胞治疗产品的生产将一直是颇具挑战性的基本问题。

7.1.3　人体细胞和移植物的临床治疗应用

细胞治疗是利用人体细胞作为药剂来缓减病情。必须注意，细胞治疗并非新技术。其实输血就是一种细胞治疗，它已在临床上应用了数十年，并获得了极大的成功

和收益。输血治疗可以将红细胞（Red Blood Cell，RBC）作为植入物输入贫血病人体内，红细胞可以储存和运输充足的氧气。同样，血小板也可以成功地输入病人体内，用于治疗血液凝固性疾病。而骨髓移植（Bone Marrow Transplantation，BMT）也已经实践了近 20 年，成千上万的癌症患者在接受高剂量化疗和放疗之后进行了骨髓移植。最近，造血干细胞的移植也越来越频繁地用于治疗造血功能紊乱疾病。这些细胞治疗法都与血细胞和血细胞的再生（hematopoiesis）㊀相关。由此可见，众多患者目前都已经享受到了细胞治疗带来的好处。生物体的移植有 3 种不同类型：异种移植，即供体与受体属于不同种系；同种异基因移植，即供体与受体属于同一种类但具有不同基因；同基因移植，即供体与受体具有相同的基因，如孪生子之间的移植。同基因移植也包括自体移植，即取自患者身上的细胞重新植回自己体内。历史上，同种异基因移植很常见，几乎已经用于所有人体器官的移植。不过，随着细胞培养和细胞操作技术的发展，自体移植也变得越来越普遍。另外，有些实验室正在试图开发“万能供体”细胞源和细胞系。这种细胞源应该具有最少或者根本没有免疫反应问题，就像 Rh 阴性的 O 型血一样㊁。

作为组织工程领域的主要研究方向之一，体外构建生物组织以及批量生产临床可用的细胞具有广泛的应用前景。表 7.1 归纳了美国每年各种人体组织和器官损伤手术的需求量。虽然手术量并不是非常大，但是这些手术每年的总费用却高达 4000 亿美元，因此，细胞治疗的社会经济效益非常大。进一步的研究将不断开发新型医疗产品，改进治疗方法，降低治疗费用，从而大大改善病人的生活质量，并重建他们的社会活动能力。

表 7.1　美国各种人体器官和组织损伤的年手术量①

适　应　证	年手术量
皮肤	
烧伤②	2150000
压伤	150000
静脉淤积性溃疡	500000
糖尿病性溃疡	600000
神经肌肉疾病	200000
脊髓和神经系统疾病	40000
骨骼	
关节置换	558200
骨移植	275000
内部固定	480000
面骨重建	30000
软骨	
膝关节表面置换	216000
髌骨软骨软化症	103400
半月板修复	250000
膝关节炎	149900
髋关节炎	219300
指关节和小关节	179000
分离性骨软骨炎	14500
肌腱修复	33000
韧带修复	90000
血管系统	
心脏	754000
大小血管	606000
肝脏	
代谢疾病	5000
肝硬化	175000
肝癌	25000
胰腺（糖尿病）	728000
肠道	100000
肾脏	600000
膀胱	57200
输尿管	30000
尿道	51900
疝	290000
乳房	261000
输血	18000000
牙科手术	10000000

① 摘自 Langer 和 Vacanti，1993。
② 烧伤病人中每年住院治疗的有 150000 人，死亡的有 10000 人。

㊀ 英文的“hemato”源于希腊文，意思就是“血液”，而“poiesis”的意思是生长。——原文注
㊁ 文中指的是美国人。西方人的 Rh 血型多为阴性，而中国汉族人的 Rh 血型多为阳性（占 99%以上）。——译者注

冯元桢教授于1985年最早提出了将生物组织直接工程化的理念，1988年Richard Skalak和Fred Fox组织召开了该领域的第一次学术会议。从此以后组织工程领域飞速发展，出现了众多学术论文，并出版了《组织工程学》（Tissue Engineering）学术期刊。目前先进的细胞治疗技术包括各种免疫疗法，以及用软骨细胞修复软骨损伤，用肝细胞和肾细胞进行体外辅助治疗，用β胰岛细胞治疗糖尿病，用皮肤细胞治疗溃疡或烧伤病人，用基因修饰成肌细胞治疗肌营养不良症等各种技术。另外，细胞治疗还包括应用组织工程植入体替代血管、输尿管、尿道等损伤或缺损的组织。各种不同组织的细胞治疗所面临的问题各不相同，下面几节将分别举例说明。

1. 骨髓移植

如图7.3所示，骨髓是人体中增殖能力最强的器官，它每天要产生4000亿数量级的骨髓细胞，这些细胞都来自少量的多能干细胞。骨髓共含有5000～10000亿个细胞，每2～3天自行更新一遍，以保持正常的造血功能。在异常的造血需求情况下，如出现全身性感染或镰刀型红细胞贫血症时，人体血细胞的生成速度还会超过正常水平。骨髓细胞的增殖能力使其特别容易受到放疗和化疗的损伤，因此，骨髓的损伤会限制这些治疗方法的继续使用。超剂量的放疗和化疗还会彻底摧毁骨髓（称为清髓），此时，如果没有其他造血支持，病人就会因丧失造血功能而死亡。

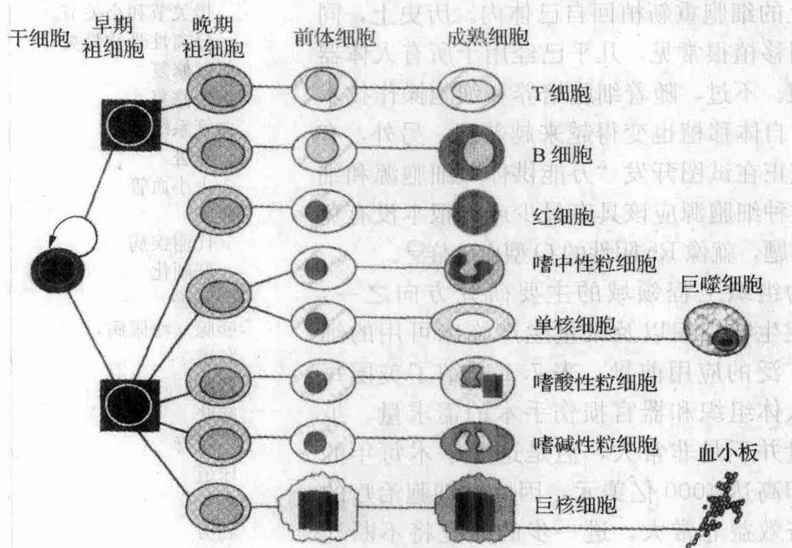


图7.3 造血细胞的产物

注：根据血液循环系统中各种细胞的稳态浓度、总血容量以及各种细胞的半衰期，可以计算整个产物谱系中各种细胞的产生速率。注意，每天4000亿个细胞的产量其实只是来自少量的干细胞。（摘自Koller和Palsson，1993）

开发骨髓移植方法就是为了解决这个问题。自体骨髓移植治疗时，在放疗或化疗之前先从病人身上采集骨髓并低温保存。等到病人使用了化疗药物并经过几个半衰期之后，再将骨髓快速解冻，重新植入病人体内。骨髓细胞直接注入血液循环系

统，而骨髓干细胞则送回骨髓腔，用于重建骨髓功能，这些细胞可以在体内重建造血组织。整个自体移植过程需要持续数周，在此期间，病人处于免疫功能受损状态。

自体骨髓移植的细胞治疗方法只是采集病人的骨髓细胞并暂时保存在体外。随着体外造血细胞培养技术的发展，其他新的治疗方法正在不断出现。其实，骨髓干细胞的采集也已经有许多新方法。其中有些方法先利用细胞因子或细胞毒素促使干细胞进入血液循环系统，然后再从循环血液中采集造血干细胞和祖细胞，并去除白细胞。

清髓性同种异基因骨髓移植常用于白血病治疗，治疗时，所使用的超剂量药物在去除疾病的同时也摧毁了病人原有的骨髓。然后，让供体细胞重新在患者骨髓中安“家”并增殖，就像自体移植一样。同种异基因移植的主要难题是死亡率较高，有 10%～15%。死亡主要由移植物与宿主机体之间的对抗反应引起，也就是，移植的细胞会将受体组织视为异物而进行攻击。如果能够克服这种排斥作用，那么，同种异基因移植技术就可以快速发展。

骨髓移植是发展较成熟且已被广泛接受的一种细胞治疗技术，它已用于多种适应证，包括治疗白血病等疾病的同种异基因移植以及治疗淋巴瘤、乳腺癌和睾丸癌的自体移植等。自 20 世纪 80 年代中期以来，骨髓移植的应用得到了快速发展，1995 年全世界大约有两万名病人接受了这种治疗，其中自体移植和同种异基因移植大约各占一半。

2. 皮肤移植体和血管移植体

皮肤是人体第三大增殖组织，如图 7.4 所示，它大致分为两层：真皮，它主要由成纤维细胞组成；表皮，它主要由处于不同分化时期的表皮细胞构成，表皮细胞

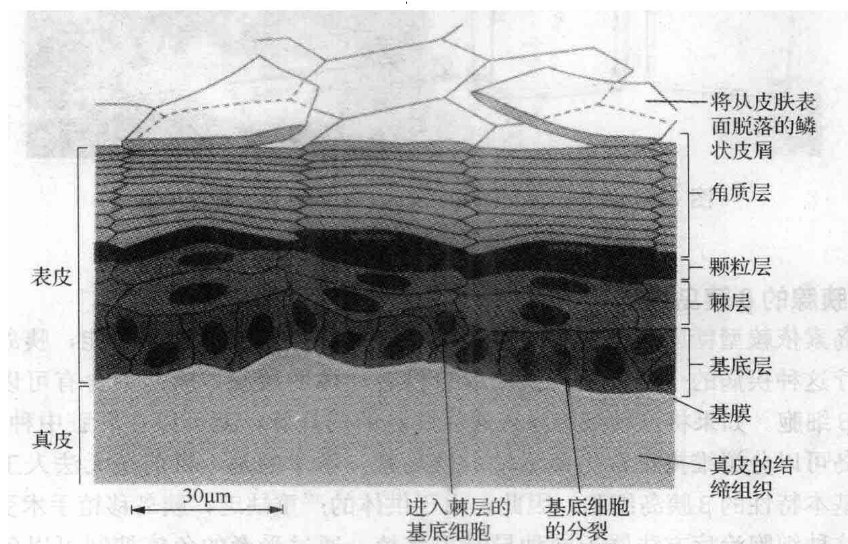


图 7.4　皮肤的细胞组成及其分化

注：图中显示了皮肤横截面上表皮细胞的排列以及处于不同分化时期的各种细胞（摘自 Alberts 等，1994）。

最终分化成为角化细胞。成纤维细胞和表皮细胞都很容易培养，它们的体外培养已经不成问题。有趣的是，真皮成纤维细胞的移植是非免疫性的，不会引起免疫反应。

皮肤移植技术常用于皮肤严重烧伤的病人或者糖尿病性皮肤溃疡病人。治疗时，先在体外培养皮肤组织，然后用于受损部位。各种皮肤细胞治疗技术已经相当成熟，不过，皮肤移植体的商品化仍然存在一些问题，医用人体皮肤替代品还不是那么容易获得。图 7.5 所示是 Advanced Tissue Sciences 公司研发的构建皮肤移植物的生物扩增反应器，它可以将人体阴茎包皮组织培养成皮肤移植体。反应器由铰链连接的两个灌流培养室组成（左图）。其移植物产品很像无毛囊的自体移植皮肤，很容易从反应器中取出（右图）。但是，这种皮肤移植体的费用很昂贵，这是限制其推广应用的主要原因。根据美国食品与药品管理局（FDA）的评估，包皮组织供体的售价过高，且所需面积过大。另外，FDA 已批准的培养基价格也很高，而且使用的每种成分都需要检测记录，所以，其利润很低。并且，一旦包皮培养被污染，由此造成的损失就是上百万美元。除了皮肤移植体之外，尽管困难重重，有人还是利用生物可降解的小管内培养的间充质干细胞，制造出了血管移植体。这些基础研究成果都还有待于临床应用的检验。

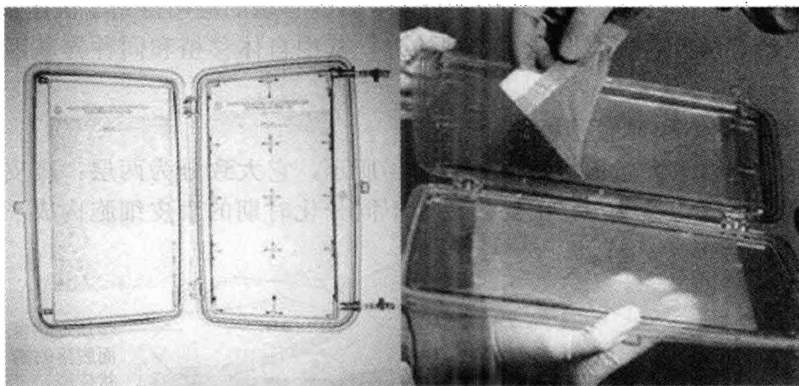


图 7.5 Advanced Tissue Sciences 公司的生物反应器
（它可以利用人体包皮组织培养人体皮肤产品）

3. 胰腺的 β 胰岛细胞

胰岛素依赖型糖尿病患者的胰腺丧失了合成并分泌胰岛素的功能，胰岛细胞移植是治疗这种疾病的一种方法。胰岛取自供者尸体的胰腺，胰岛中含有可以分泌胰岛素的 β 细胞。如果将这种细胞注入通往肝脏的门静脉，就可以在肝脏中种植胰岛，这种胰岛可以分泌维持正常生命活动的胰岛素。不幸的是，目前还无法人工培养能够保留基本特性的 β 胰岛细胞，因此，由于供体的严重缺乏，胰岛移植手术受到很大限制。这种细胞治疗方法属于同种异基因移植，通过受者的免疫抑制可以延长移植体的存活时间。

细胞治疗的免疫排斥是一个重大问题，本章后面会单独讲述这个问题。如图 7.6

所示，克服排斥反应的一种方法就是物理隔离，也就是让植入体与受体之间通过一层半透膜进行物质交换，避免供体细胞与受体免疫系统的直接接触。

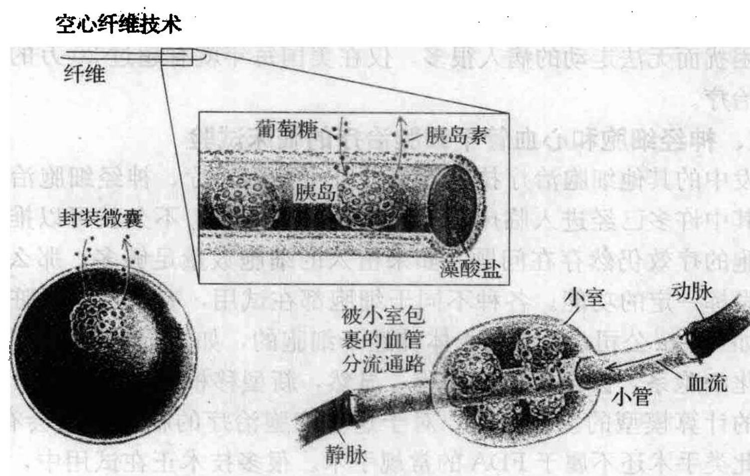


图 7.6　封装在塑料半透膜中的胰岛可以免受免疫系统的攻击，这种方法有很好的应用前景（摘自 Lacey，1995）

4. 软骨与软骨细胞

软骨是一种特殊的组织，它无血管、无淋巴、也无神经。它的主要成分是细胞外基质，其中散布着密度很低的软骨细胞，每毫升大约只有 100 万个细胞，而一般组织中的细胞密度要比这个数值大几百倍。软骨细胞可以通过体外培养增殖约 10 倍。利用这种技术，膝关节深部的软骨损伤可以通过自体移植的方法治疗。如图 7.7 所示，治疗时，首先在病人膝关节受损区域以外采集健康的软骨活组织；然后，经

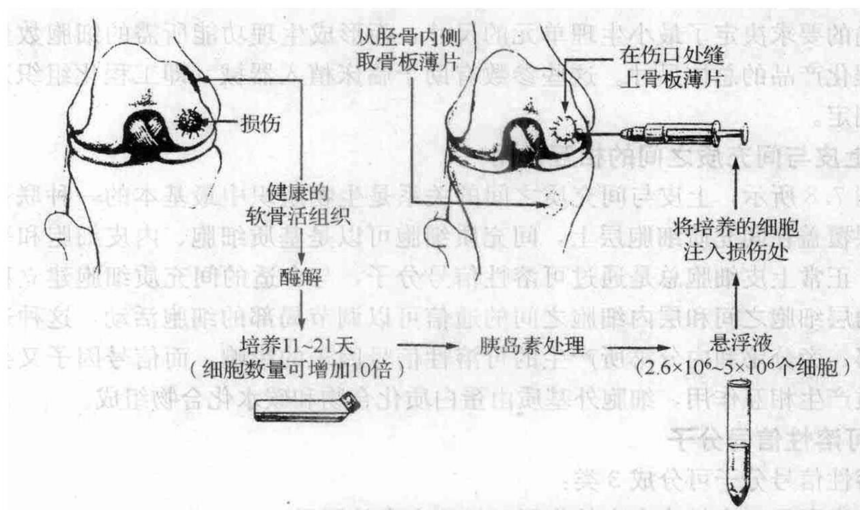


图 7.7　右股骨头部软骨细胞移植技术示意图（摘自 Brittberg，1994）

过酶解处理从基质中分离出软骨细胞，再经过两个阶段的细胞培养；最后，就可以从培养基中提取出细胞，并植入膝关节的受损部位。这样，移植的软骨细胞就可以愈合损伤部位。这种手术的原理似乎比较简单，但其临床实现并不那么容易。备受膝关节疾病困扰而无法走动的病人很多，仅在美国每年就有超过 20 万的病人需要进行这种细胞治疗。

5. 肝脏、神经细胞和心血管干细胞治疗的临床试验

正在开发中的其他细胞治疗技术还有肝脏干细胞治疗、神经细胞治疗和心血管细胞治疗，其中许多已经进入临床试验阶段并即将完成，不久就可以推广使用。不过，植入细胞的疗效仍然存在问题。如果植入的细胞数量足够多，那么，它们会生长起来，并发挥一定的功能。各种不同干细胞都在试用，有从尸体肝脏中提取成人干细胞的，如 Vesta 公司；也有用人体神经干细胞的，如 Stem Cell 公司；还有其他公司使用转化细胞系，甚至胚胎干细胞。显然，新型移植硬件装置的开发，以及有关物质传输的计算模型的预测分析，对于这些细胞治疗的临床应用会有很大帮助。不过，目前此类手术还不属于 FDA 的常规手术。很多技术正在试用中，而有关物质传输的模型分析工作却很少进行。其实，有些问题必须得到解答，例如：为了产生组织功能并维持细胞的存活，细胞团的最大直径应该是多少？要使细胞治疗在体内发挥疗效，最佳物质灌输速率应该为多少？

7.1.4 组织调控机制

工程化的组织应该具备正常生物组织的功能，因此，在生产细胞用于构建组织时要以自然生理规律为指导。生物组织的动态调控机制主要有两种：一是上皮与间充质之间的相互作用；二是干细胞与成熟细胞系之间的联系。

了解生物组织的生理机制之后，我们就可以集中精力解决工程化的相关问题。物质传输的要求决定了最小生理单元的尺寸，而形成生理功能所需的细胞数量则决定了工程化产品的总体尺寸。这些参数有助于临床植入器械（即工程化组织）总体目标的制定。

1. 上皮与间充质之间的相互作用

如图 7.8 所示，上皮与间充质之间的关系是生物组织中最基本的一种联系。上皮细胞层覆盖在间充质细胞层上，间充质细胞可以是基质细胞、内皮细胞和平滑肌细胞等。正常上皮细胞总是通过可溶性信号分子，与合适的间充质细胞建立稳定的联系。两层细胞之间和层内细胞之间的通信可以调节局部的细胞活动，这种通信受到自分泌、旁分泌和内分泌所产生的可溶性信号因子的影响，而信号因子又会与细胞外基质产生相互作用，细胞外基质由蛋白质化合物和碳水化合物组成。

2. 可溶性信号分子

可溶性信号分子可分成 3 类：

自分泌因子：由细胞产生并作用于细胞自身的因子；

旁分泌因子：由细胞产生并作用于附近其他细胞的因子；

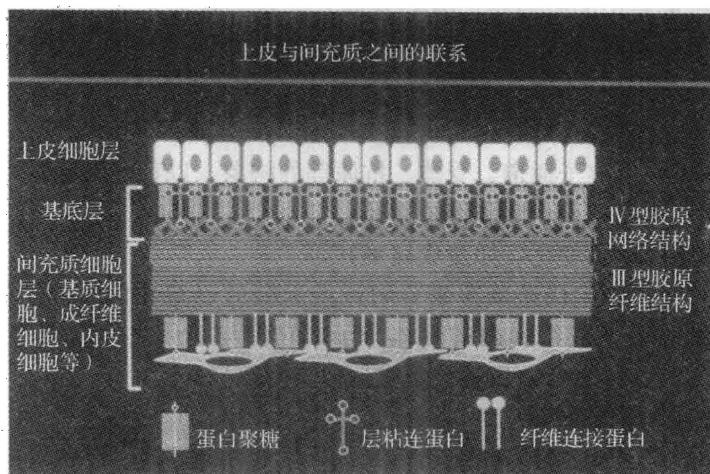


图7.8 上皮与间充质之间的联系

内分泌因子：是全身因子，由离开靶细胞一定距离的细胞产生，通过血液或淋巴液输运到靶细胞部位。

自分泌和旁分泌的信号因子有很多种，根据细胞外基质的化学成分，它们可以促进细胞分裂（即引导细胞生长），也可以驱使细胞分化（即诱导组织特异性功能的表达）。其中已经研究得比较充分的因子有：胰岛素样生长因子（IGF）、表皮生长因子（EGF）、成纤维细胞生长因子（FGF）、集落刺激因子（CSF）、血小板源性生长因子（PDGF）、肿瘤生长因子（TGF）以及白细胞介素等细胞因子。多数因子都被称为"生长因子"，但千万不要误解，其实，这只是因为这些因子最初是在科学家们研究某类细胞的生成过程中被发现的。

3. 细胞外基质

细胞外基质（Extra Cellular Matrix，ECM）可以分成两种，一种存在于彼此相连的同类细胞之间，如上皮细胞与上皮细胞之间；另一种存在于不同种类细胞之间，如上皮细胞与间充质细胞之间。其中位于不同种类细胞之间的细胞外基质又称为基底膜。细胞外基质所包含的蛋白质占人体总量的1/2以上。胶原蛋白是细胞外基质的主要成分之一，它由25种以上不同分子构成，约占人体蛋白质的1/3。过去很长一段时间里，人们一直以为细胞外基质只是起着机械支撑作用，它将各个细胞按特定的方式结合在一起。但是，近20年以来，人们发现细胞外基质其实是人体的一个重要信号系统，它处于固态形式，并且发送慢性持续性信号。同类细胞之间的ECM的两种主要成分是：一是具有组织特异性和年龄特异性的细胞粘附分子（Cell Adhesion Molecules，CAM）；二是蛋白聚糖，其分子以蛋白为核心，周围粘附了带负电荷的硫酸化多糖（即糖胺聚糖），糖胺聚糖有硫酸乙酰肝素、肝素、硫酸软骨素和硫酸皮肤素等不同形式。不同种类细胞之间的基底膜ECM含有层粘连蛋白和纤维连接蛋白等粘附分子，它们通过基质受体（即整合素）将细胞与一种或多种胶原支架联合

在一起。相邻细胞层中的胶原蛋白互相交联在一起，形成了各个细胞层之间的稳定连接。而且，蛋白聚糖可以与粘附分子、胶原蛋白和基底细胞表面键合在一起。

两种 ECM 的组成成分会向细胞直接发送慢性持续的信号，并且它们还可以通过以下方式间接发送信息：

1）调整细胞膜上离子通道、受体、抗原等的构象，以稳定细胞；

2）影响细胞内的各种代谢和调控途径；

3）诱导细胞的形状变化，如变成扁平状等。

这些反应可以使细胞快速响应来自附近或远处的可溶性信号分子的作用。可溶性因子作用于细胞表面或细胞内的受体分子，它们与受体之间具有很高的亲和力，一旦信号分子与其受体结合之后就会触发信号转导反应。

4. 人体各种细胞的数量

人体组织的细胞密度约为每毫升 10～30 亿个。重 70kg 的成人的体积大概是 70000ml。因此，人体的细胞总量大约为百万亿个（即 10^{14}）。人体器官的大小一般约为 100～150ml，因此，每个器官约含有 1000～15000 亿个细胞（即 $1\times10^{10}\sim15\times10^{10}$）。器官由许多功能亚单位组成（参见 7.4 节），它们的尺度一般在 100μm 数量级。边长为 100μm 的正方体中所含的细胞估计约为 500～1000 个。表 7.2 归纳了一些有关人体细胞数量级的数据。

表 7.2 组织生物学和组织工程中常用的人体细胞数量级的数据

体内细胞数量	
全身	10^{14}
人体器官	$10^{9}\sim10^{11}$
功能亚单位	$10^{2}\sim10^{3}$
体内细胞增殖的产量	
理论上单细胞的最大产量（即 Hayflick 界限）	$2^{30\sim50}<10^{15}$
人一生中骨髓的血细胞产量	10^{16}
人一生中小肠上皮细胞的产量	5×10^{14}
体外细胞增殖的产量	
一般细胞治疗所需的细胞数量	$10^{7}\sim10^{9}$
人体组织的扩增潜能①	
造血细胞	
单核细胞	10 倍
CD34 富集细胞	100 倍
抗原富集	$10^{6}\sim10^{7}$倍
T 细胞	$10^{3}\sim10^{4}$倍
软骨细胞	10～20 倍
肌细胞、真皮成纤维细胞	$>10^{6}$倍

① 扩增潜能是指体外培养时单个细胞能够生产的细胞数量。

由此可见，器官一般含有几亿个功能亚单位，这个数量是由每个亚单位的性能和人体生理功能的需求决定的。例如，肾脏中肾单位的数量就是由人体排泄毒性产物所需的最大清除率以及单个肾单位的清除能力所决定的。一般人体组织的基本功

能亚单位只包含几百个细胞，而且是混合细胞群体。正如 7.3 节将阐述的那样，多数器官含有辅助细胞，可占细胞总数的 30%。并且，如下文所述，组织特异性细胞可能处于细胞分化的许多不同时期。

为了能够精确地重建生物组织的功能，必须了解组织微环境的性质，以及细胞动力学、细胞通信和各种代谢过程。要生产能够用于临床治疗的细胞数量需要大量微环境，而且，这些微环境必须与各种功能亚单位的运作环境相似。因此，用于细胞培养的装置必须能够均匀地提供营养物质、氧气和生长因子等物质，并且，在 100μm尺度以上这些物质的浓度必须均匀一致，而小于这个尺度的微环境中应该存在非均一性，这是组织功能亚单位正常运作所必须满足的条件。

另外，细胞移植的方法也是需要考虑的问题，将体外生成的组织植入到受者体内会引起很多问题。目前还不太可能在体外人工合成可用于移植的功能完整的器官。现在已经可以施行的方法是生产并移植干细胞和祖细胞，使它们在体内能够充分增殖，从而重建器官功能。骨髓移植使用的就是这种方法。

7.1.5　临床应用的若干问题

1. 具有临床应用价值的细胞数量是多少?

见表 7.2，目前，临床实践和实验研究中细胞治疗所需的细胞数量在“几千万”到“几十亿”之间。由于组织培养所需的细胞密度高于每毫升 1000 万个，因此，用于细胞治疗的细胞培养装置的体积大约应该为数十到数百毫升。

2. 正常细胞的增殖有哪些基本限制?

单个细胞能够经历的分裂次数由其所处的发育阶段决定，处于不同发育阶段的细胞主要有 3 类：二倍体的多能干细胞、二倍体的单能体细胞亚群以及多倍体的体细胞亚群。其中，干细胞可以自我复制，并且可以经历无数次分裂，它们是各种永生细胞系和肿瘤细胞的潜在来源。二倍体的体细胞的分裂次数受到所谓的 Hayflick 界限的限制。正常人体的二倍体细胞在体外培养时大约能够分裂 30～50 次，因此，理论上，单个二倍体的体细胞可以培养出 10^{10}～10^{15} 个细胞。对于细胞治疗的要求，无论是干细胞还是体细胞，Hayflick 界限都不是这些二倍体细胞增殖的苛刻条件。另外，各种组织中存在的多倍体细胞亚群的数量比较少，其扩增潜能也很小，几乎可以忽略不计。不过，肝脏和心脏等的休眠组织中存在的多倍体细胞比较多。如果细胞在再生信号的刺激下合成了 DNA，但却没能够完成细胞分裂，那么，就会形成多倍体细胞。这种细胞的体积较大，因此也被称为肥大细胞。于是，组织的再生活动既包括了细胞增生（即二倍体细胞亚群数量的增加）也包括细胞肥大（即多倍体细胞亚群的增加）。

3. 正常细胞体外培养时的生长速度有多快?

各种正常细胞体外培养的生长速度各不相同。造血祖细胞的倍增时间约为 11～12h，它是人体成年细胞的最快增殖周期。包皮真皮层的纤维细胞的倍增时间约为 15h，其生长速度较快，部分原因可能是因为这种细胞来自干细胞和祖细胞比例很高

的新生组织。成年软骨细胞体外培养的生长速度较慢，其倍增时间约为24～48h。

4. 目前这些细胞是如何生产的？

细胞扩增在各种培养器皿中进行，有T形瓶、转瓶以及用于T细胞的培养袋、用于造血细胞等细胞系的悬浮细胞培养瓶等。新型支架材料也可以用于多数细胞的体外扩增。如果要达到各种特定种类细胞的最大扩增潜能，那么就需要精确设定其培养条件，包括细胞外基质的特殊形式、相应的激素和生长因子混合物、营养物质以及包含特定浓度钙离子、微量元素和气体分子的基质。最近出版的几本教材中都提供了各种细胞培养的条件，请参考本章末尾的推荐阅读资料。

7.2 生物学问题

7.2.1 干细胞

各种组织中都含有干细胞池，用于细胞增殖，增生的细胞成熟后就表现出各种组织的表型。干细胞的成熟过程包括定型阶段和分化阶段。定型阶段中多能干细胞产生子细胞，这些子细胞具有单能细胞的遗传潜势；分化阶段细胞中的某几套基因被激活，产生各种不同的表达水平。为了解释如何将干细胞用于组织功能的修复，以下几小节将详细介绍干细胞的分化过程。

1. 各种干细胞

多能干细胞能够产生子细胞，而这些子细胞具有多种发育命运，能够进行自我复制，并且能够再生成与母细胞相同的子细胞。全能干细胞能够产生生物体所有不同种类的细胞。而定向干细胞的遗传潜势则被限定在几种发育命运上，它们只能发育成为若干种类型的细胞，并不能生成机体所有不同种类的细胞。例如，皮肤干细胞只能产生皮肤中的各种细胞，而不能产生心脏中的细胞；肝脏干细胞则可以产生各种肝脏细胞，而不能产生脑细胞。某些出版物把定向干细胞称为“成年干细胞”，这并不合适，因为定向干细胞其实在胚胎和成人组织中都存在。定向干细胞可以生成单能祖细胞，也称为定向祖细胞，其遗传潜势只限定于一种细胞命运。这种单能祖细胞能够迅速增殖，产生大量细胞，这些细胞随即分化为成熟细胞。各种干细胞和单能祖细胞都属于正常细胞，而肿瘤细胞和永生细胞系却是异常细胞。目前已经识别的定向干细胞的体积都很小，直径一般为6～10μm，但其核质比却很大[⊖]，并且具有某些早期基因的表达（如甲胎蛋白）和抗原的表达（如CD34、CD117等）。定向干细胞染色质能够结合的特定染料分子的浓度要比成熟细胞低，因此，可以利用流式细胞仪技术将其分离出来。干细胞表达的端粒酶能够使其染色体的端粒保持恒定的长度，从而使干细胞在体内或者体外能够无限制地进行分裂。任意一种定向干细胞的分离和纯化都需要用到多个参量，如抗原类型、细胞大小、细胞密度等。没

⊖ 核质比是指细胞核与细胞质的容积之比。——译者注

有一个参量能够单独用于定义定向干细胞。而且，定向干细胞在体内的生长速度很慢，还可能随机定型并分化。造血干细胞的倍增时间约为6周，该周期是通过慢速拍摄技术测量得到的。体外培养时，造血干细胞可以定型并分化，第1、2次倍增大约需要60h，随后其周期会缩短到24h，到第5、6次倍增时，其最大速度可以达到每12～14h倍增一次。

2. 干细胞存在的证据

小鼠接受致死剂量的射线辐射之后会由于造血功能的丧失而死亡，但是，只要使用20种选定的干细胞就可以挽救这种小鼠的生命。正如干细胞模型所预测的那样，这些动物可以重建多种造血细胞系。而且，接受了亚致死剂量的动物，经过一段时间之后，其骨髓中的间充质干细胞会生成各种器官组织的细胞。这些研究结果与其他许多研究一起都毫无疑问地证实了干细胞的存在以及它们的多能性，而且表明干细胞在体内可以长期存在。

3. 干细胞的微环境

干细胞微环境的研究是一个相当新的领域，它研究干细胞池的局部微环境。这个领域太新了，在未来几年内很可能会发生巨大变化。如下这些已经归纳出来的知识对于体外干细胞扩增条件的设计具有重要的指导意义。

1）干细胞不能通过酶反应机制由单一脂质源产生所需的各种脂质衍生物，因此，它们需要复合脂类混合物的支持才能生存并维持正常生理功能。

2）在决定细胞扩增还是分化时，钙离子浓度具有非常重要的作用。这种现象的机制还不清楚。

3）铜等特殊微量元素可以加快某些定向干细胞的分化。目前还不知道这种作用是否适用于所有干细胞种类，也不清楚其中的机制。

4）干细胞微环境必须含有各种特殊的激素和生长因子，其中最常见的是胰岛素和转铁蛋白。增加其他因子还可以导致定型祖细胞的扩增，使得干细胞向特定的细胞命运发育。

据了解，干细胞池的基质中包含有年龄特异性和组织特异性的细胞粘附分子、层粘连蛋白、胚胎胶原蛋白（如Ⅲ型和Ⅳ型胶原）、透明质酸以及胚胎蛋白聚糖等物质。当干细胞逐渐向特定的细胞命运发育成熟时，基质的化学成分也要以一定的梯度形式向某种成熟细胞的基质转变。虽然每种成熟细胞胞外基质的化学成分是特定的，但它们都包含年龄特异性细胞粘附分子、各种纤维胶原蛋白（如Ⅰ型和Ⅱ型胶原）、纤维连接蛋白以及年龄特异性蛋白聚糖等。对于包皮细胞和神经细胞，由于其成熟细胞没有胶原蛋白、纤维连接蛋白和层粘连蛋白的表达，因此，这些成熟细胞基质的化学成分以细胞粘附分子和蛋白聚糖为主。

干细胞的活动取决于基质中的各种年龄特异性和组织特异性的信号分子，其中包括白血病抑制因子（LIF）、各种成纤维细胞生长因子（FGF）以及IL8和IL11等白细胞介素等。基质的信号分子系统也还没有完全弄清楚。

干细胞微环境信号系统中最缺乏了解的应该是反馈作用信号。这种信号来自成

熟细胞，对干细胞的增殖具有抑制作用。反馈作用的存在似乎表明体外细胞扩增时需要将具有分裂能力的细胞（即二倍体亚群）与成熟的非增殖细胞隔离开来。

4. 含有干细胞的人体组织

以前数十年里，人们一直认为干细胞池只存在于皮肤、骨髓和肠道等可以快速增殖的组织中。现在，越来越多的证据表明所有人体组织中都含有干细胞池，甚至中枢神经系统也含有干细胞（Tramontin 等，2003）。

5. 干细胞的作用

干细胞池是组织发生转变和再生的根本资源。干细胞的定型可以诱发细胞的替换和组织的生成，并由此实现组织的修复和组织功能的维护。如果疾病或药物等毒素引起干细胞的缺失，那么，最终就会导致人体器官部分或者完全丧失功能。另外，干细胞的基因突变会导致肿瘤的产生，因此，实际上存在正常和异常两种干细胞。于是，现在有人把肿瘤看作“转化干细胞”，干细胞生物学家已经证实了这种假设（Galli 等，2004）。

7.2.2 成熟细胞系的生物学基础

所有干细胞都是多能的，可以生成多种不同细胞系的子细胞，这些子细胞再逐步分化，变成各种组织的成熟细胞。图 7.9 所示是描述各种组织特异性干细胞产生成熟细胞的一种通用模型。多能的定向干细胞在体内复制速度缓慢，并且，其复制速度受到各种全身性信号分子的调节。复制所产生的直接子代细胞被定型成为单能祖细胞，这种祖细胞能够快速增殖并形成组织（如皮肤等）。这是组织对于温和的正常再生性刺激的急性响应过程，也就是，单能祖细胞经过多个中间过渡阶段逐步变为完全成熟的细胞。其实，各种不同的组织特异性功能是由整个成熟过程细胞系中的各种细胞共同形成的。通常，根据逐渐变化的表型，可以将祖细胞分为“早期”、

	干细胞	早期祖细胞	晚期祖细胞	前体细胞	成熟细胞
细胞数量	每个细胞的增殖潜能为2^{30}~2^{50}				一生的总需求量约为10^{16}
细胞周期	很长（t_d≈6周）	长（t_d≈60~100小时）	短（t_d≈12小时）	长	无（某些特殊情况下可以被激活）
细胞凋亡	无活性	无活性	活性很高（1/5000存活率）	活性很低	无活性（可被诱导）
细胞运动	无（除了回归原位）	无	缓慢	较快	与生理功能相关
调控方式	细胞之间接触	细胞之间接触	可溶性生长因子	可溶性生长因子	可溶性生长因子

图 7.9 增殖组织的细胞生产模型

注：该模型从几十年血液学研究结果中归纳得到。图中每一列表示逐渐分化成熟的各种细胞，每一行则表示细胞的命运和活动等情况（见图 7.3），t_d是倍增时间。

"中期"、"晚期" 3类。有些特殊基因只在某个特定的时期才表达；有些基因则在整个细胞系中都有同工型基因表达；而另一些基因在各个时期的表达水平各不相同。最后，细胞将进入衰老过程，也称为细胞凋亡，这是细胞衰老的特征现象。细胞的成熟过程一方面与染色质的变化等细胞固有机制相关，另一方面还与基质及其微环境中的可溶性信号分子有关。微环境包含了影响细胞生长、分化和凋亡的信号分子。例如，某些生长因子属于生存因子，具有抗衰老作用。

干细胞与成熟细胞系的模型举例

研究较深入的干细胞模型有造血干细胞、皮肤干细胞和肠道干细胞。

1. 骨髓和血细胞的形成

造血干细胞是最早建立的干细胞模型（见图7.3）。小鼠干细胞移植实验已经证明，只要用少得惊人的细胞数量就可以重建造血细胞的多种细胞系。一旦细胞到达某个成熟期，它们就会离开骨髓，进入血液循环系统，并行使其成熟细胞的功能。成熟细胞最终会死亡并被新的细胞替代。成熟细胞的死亡率（包括凋亡和坏死）决定了组织的细胞生产速度，也就是决定了所需的定型干细胞数量。图7.3显示了造血细胞系的特殊产物。

2. 小肠绒毛

如图7.10所示，小肠内壁由吸收营养物质的绒毛构成。小肠上皮细胞层具有高度的动态性，其细胞大约每5天就要更换一次，因此，小肠是人体第二大增殖组织。

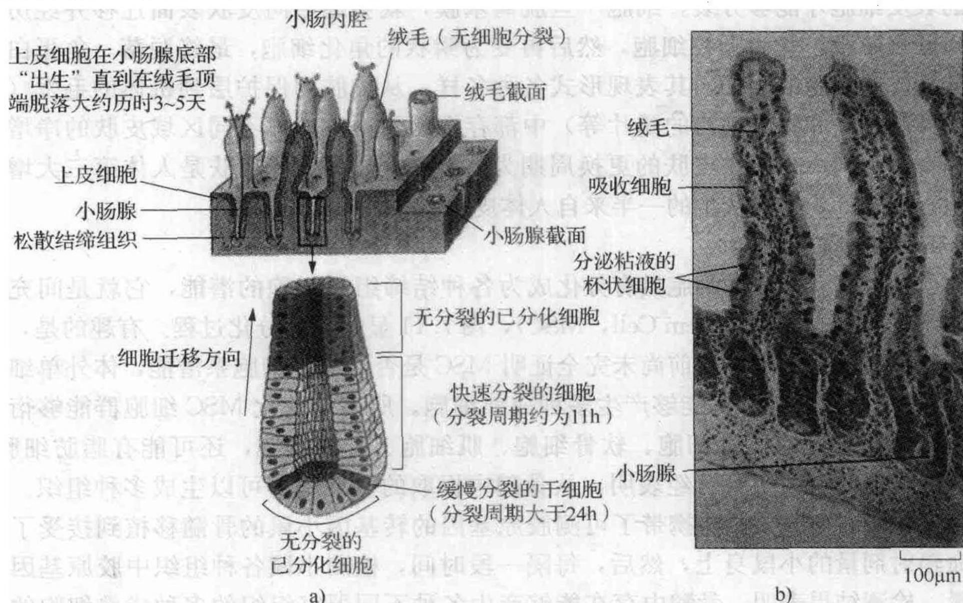


图7.10　小肠绒毛（摘自Alberts等，1994）

a）绒毛和小肠腺示意图，以及小肠腺各部位细胞的分裂状态

b）绒毛上的各种小肠上皮细胞，绒毛的直径约为80μm

小肠绒毛之间存在管状的上皮凹褶，这些凹褶就是小肠腺。所有小肠上皮细胞都在小肠腺的某个部位产生，一旦上皮细胞成熟，它们就迁移到小肠腺外缘，然后，大约经过 5 天时间的迁移运动，这些细胞会从绒毛的底部移动到顶端，并在绒毛顶端凋亡和脱落。在此过程中，它们作为成熟的实质细胞完成组织特异性功能，也就是吸收和消化小肠内腔中的营养物质。每个小肠腺的底部都有一圈缓慢分裂的干细胞，数量约为 20 个。干细胞分裂产生的子细胞会移动到小肠腺上部，然后变成快速增殖的祖细胞，其倍增时间约为 12h。这些细胞在移动过程中不断分化，当它们离开小肠腺并进入小肠绒毛的底部时已经成熟。

3. 皮肤

如图 7.4 所示，人体皮肤主要分为表皮和真皮两层，两层细胞之间由一层细胞外基质隔开，称为基膜。Ⅶ型胶原蛋白是基膜的重要成分。两层细胞之间的基膜起伏不平，构成了深浅不同的囊袋结构。人体的躯体、手掌、面部等不同部位的皮肤，其表皮层和真皮层之间存在独特的结构。真皮层是位于基膜下的结缔组织层，主要由成纤维细胞组成。表皮层则由多层上皮细胞组成，这些上皮细胞是处于不同分化时期的角质形成细胞（keratinocyte)，其中位于基膜上的角质形成细胞分化程度最轻。表皮最上层是鳞状柱形结构，柱形的直径约为 30μm。表皮与真皮之间形成的较深的囊袋中存在干细胞，毛囊附近的乳头中也发现有干细胞。干细胞产生定向祖细胞，这些祖细胞会迁移到较浅的囊袋中，并由此进入整个基膜层。只有附着在基膜上的表皮细胞才能够分裂，细胞一旦脱离基膜，就会逐渐向皮肤表面迁移并经历一系列分化阶段，变为颗粒细胞，然后再变为鳞状的角化细胞，最终脱落。角蛋白是已分化皮肤细胞的表征，其表现形式各种各样，从皮肤的保护层到机体的毛发（包括鸟类的羽毛和脊椎动物的鳞片等）中都存在角蛋白。人体不同区域皮肤的净增殖速率有所不同。一般，皮肤的更换周期为几个星期，因此，皮肤是人体第三大增殖组织。据估计，室内灰尘的一半来自人体皮肤。

4. 间充质干细胞

已经证实，有一种细胞具有分化成为各种结缔组织细胞的潜能，它就是间充质干细胞（Mesenchymal Stem Cell，MSC)，图 7.11 显示了其分化过程。有趣的是，这种细胞存在于骨髓中。目前尚未完全证明 MSC 是否具有多细胞系潜能，体外单细胞培养实验没有表明 MSC 能够产生多种成熟细胞。所谓的纯化 MSC 细胞群能够衍生出来的成熟细胞有：骨细胞、软骨细胞、肌细胞、纤维细胞，还可能有脂肪细胞。转基因小鼠的有关实验已经表明，从骨髓中提取的亚细胞群可以生成多种组织。在这种实验中，研究人员将携带了可测胶原基因的转基因小鼠的骨髓移植到接受了亚致命辐射剂量的小鼠身上，然后，每隔一段时间，检测小鼠各种组织中胶原基因的含量。检测结果表明，骨髓中存在能够产生各种不同器官组织的多种成熟细胞的亚细胞群。该实验结果显示了多能 MSC 存在的可能性，但并不能证明其确实存在。

5. 肝脏

现在越来越多的人认为肝脏其实是一个成熟细胞谱系组成的系统，与骨髓、皮

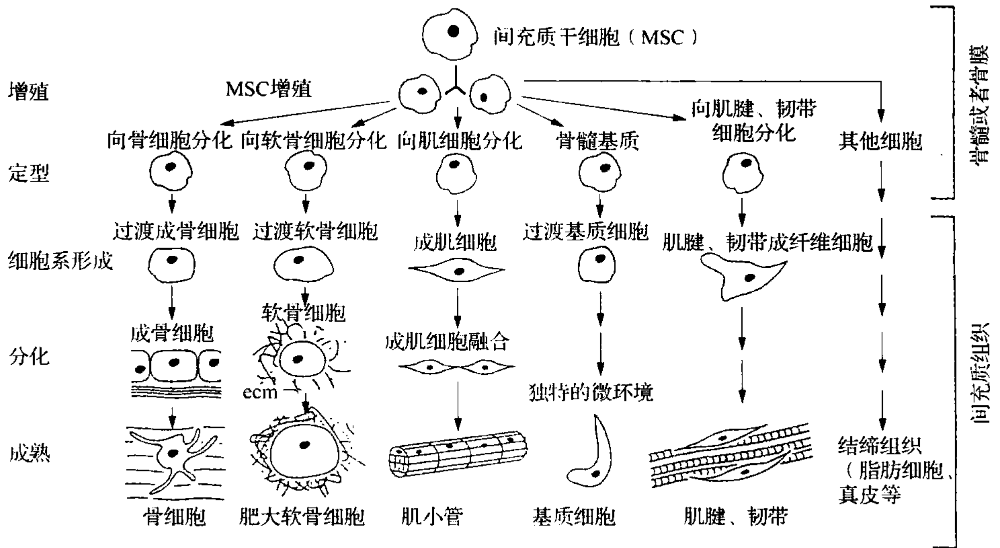


图 7.11　间充质干细胞分化成为各种不同的间充质细胞类型（摘自 Bruder 等，1994）

肤和肠道相似，它也含有干细胞池。肝脏的细胞系构成了肝腺泡的物理结构（参见本章习题 10 的图示），肝腺泡是肝脏的基本结构和功能单位，其横截面呈现为车轮形状，车轮的中轴是中央静脉，外周六个角上是门静脉、肝小动脉和胆管组成的门管三联体，而车轮的辐条则由板状排列的实质细胞构成，细胞板的两边是带窗孔的窦状内皮细胞层。传统上，肝脏被划分成 3 个环带：Ⅰ带是门管周带，Ⅱ带是肝腺泡中带，Ⅲ带是中心周带。血液从门管三联体的门静脉和肝小动脉流出，经过肝细胞板之间的窦状隙，流向中央静脉。干细胞池在门管三联体周围，也就是Ⅰ带内，在解剖学上称为 Hering 小管区。肝脏干细胞是双能的，其子细胞可以分化成为胆细胞（即胆管上皮细胞）或者肝细胞。位于不同区带的肝细胞在形态、生化特性以及功能上都有明显的差别。胆细胞可能也是如此，但它们还没有像肝细胞那样被研究得很透彻。从Ⅰ带到Ⅲ带，肝细胞的体积逐渐增大，细胞的线粒体、内质网和糖原颗粒的形态特征也随着区带的变化明显不同。而且，不同区带中肝细胞的 DNA 含量截然不同，Ⅰ带的细胞是二倍体，到了Ⅱ带和Ⅲ带则逐渐变为多倍体细胞。中心周带的八倍体细胞表现出衰老迹象。成年啮齿动物（即大鼠和小鼠）的肝脏中 90%～95%细胞是多倍体；成人肝脏中则有 40%～50%细胞是多倍体，而胚胎和新生儿的肝脏中全部都是二倍体细胞。大鼠年龄达到 3～4 周时，其肝脏细胞开始转变为成年的多倍体形式，而人则在接近 20 岁时才出现这种转变。随着年龄的增长，所有哺乳动物肝细胞中的多倍体细胞都会逐渐增加，这一点可以解释肝脏的再生能力随年龄增长而减弱的现象。

见表 7.3，肝脏不同区带所含的实质细胞在体和离体的分裂潜能各不相同，分裂能力最强的是门管周带区域的二倍体细胞，最弱的则是中心周带区域的多倍体细胞，

它们几乎不能分裂。只有二倍体实质细胞亚群能够进行完整的细胞分裂过程，这些细胞包括干细胞亚群、直径小于15μm的单能祖细胞以及平均直径为18～22μm的二倍体成年肝细胞（也称为“小肝细胞”）。不过，这些二倍体亚群的分裂能力差别很大。单个小肝细胞20天内可以生产120个子细胞，而单个干细胞或者祖细胞在同样时间和同样条件下却能够生产4000～5000个子细胞。成熟的多倍体细胞也可以进行DNA合成，但是，由于细胞分裂因子的衰弱，即使体外培养的扩增条件再好，多倍体细胞的分裂也很有限。能够进行完整分裂的细胞亚群只有二倍体细胞。虽然利用各种刺激信号可以诱导多倍体细胞进行DNA合成，但能够完成细胞分裂的却不多。实际上，通过标准的“增殖”测定方法（如BUDR、Ki67、H^3-胸腺嘧啶核苷测定等）获得的有关细胞核生长的数据既包括细胞增殖（即完整的细胞分裂过程）也包括细胞肥大（即只有DNA合成而无细胞分裂）。有关肝脏的干细胞系统近年来已经有人发表了不少综述性文章。

表7.3 不同物种的各个肝腺泡区带的肝细胞参数以及体外培养参数

参　数	物种	Ⅰ带	Ⅱ带	Ⅲ带
染色体倍数①	大鼠	2N	4N	4N和8N
	小鼠	2N和4N	4N和8N	可达32N
	人	2N	2N	2N和4N
生长能力		最强	中等	几乎为0
细胞外基质的成分		Ⅳ型和Ⅲ型胶原、层粘连蛋白、透明质酸、HSPG	从Ⅰ带到Ⅲ带按照梯度变化 ⟶	Ⅰ型和Ⅲ型胶原、纤维连接蛋白、HPPGs
基因表达		早期	中期	晚期

① 根据哺乳动物的计算数据，成熟肝细胞的直径与染色体倍数相关。二倍体细胞的直径约为18～22μm；四倍体约为22～35μm；八倍体以上的细胞直径可达75μm。

7.2.3 干细胞增殖行为的模型

至此，我们已经知道各种干细胞的复制机制对于人体器官组织功能的形成和维护以及组织工程人工移植产品的开发都至关重要。那么，干细胞是如何分裂的？它们分裂时都发生了什么呢？下面介绍描述干细胞群动态行为的3种模型。

1. 克隆演替假说

人体的细胞系统由干细胞池维持，干细胞池中的细胞要么处于缓慢生长状态，要么处于休眠状态。在组织的整个生命过程中，这种细胞池始终存在，其中的细胞受到刺激后就会进入复杂的细胞增殖和分化过程。并且，干细胞一旦被激发，单个细胞就能克隆产生大量的成熟细胞。在反馈信号的作用下，克隆细胞群体都只能维持有限的生命周期。一段时间之后，整个克隆群体就会凋亡，根据需要，新的干细胞克隆群体会接替前任的角色。

2. 确定性的自我维护与自我更新模型

该模型基于干细胞的自我复制机制，它假设：一个干细胞分裂成为 2 个子细胞之后，50%的可能性是一个子细胞保持原干细胞的特性，而另一个子细胞则进入分化过程。这个自我更新的几率是可以调控的，不一定精确为 50%，它与组织的动态变化相关。

3. 随机性模型

这种模型假设一个干细胞分裂之后，其后代子细胞中可能没有一个是干细胞（即克隆演替模型的情况），也可能有 1 个是干细胞（即确定性模型的情况），或者 2 个都是干细胞；而这 3 种情况都有一定的出现概率。

7.2.4 干细胞与组织工程

干细胞构建各种生物组织，它们是组织细胞的来源。在体内能够构建和维护各种生物组织的干细胞也可以用于体外制造生物组织，如生物人工器官。

干细胞的体外培养。体外培养干细胞，就是要让干细胞不断分裂却不能分化。理论上，这种方法可以产生用之不竭的干细胞源。不过，虽然体外培养技术发展迅速，但目前的水平还没有达到使任何干细胞家族都能够真正自我复制式地扩增。

干细胞的分离。干细胞的分离方法已有很多种，不过，目前还没有一种干细胞只要利用单一参量就可以提纯的，有效的分离方法一般都要用到多个参量。例如，利用免疫选择法可以通过干细胞表面特异性抗原识别细胞类型，再结合细胞直径、密度和粒度等参数就可以提纯干细胞。其他已发表的干细胞分离方法还有流式细胞法以及利用磁性柱、亲和柱和反向淘洗的免疫选择法等。

干细胞确认方法有体外和体内两大类。体外用集落形成法，也就是在精确控制的条件下培养单个细胞，使其扩增，然后观察其生成的子细胞是否具有多种发育命运。体内方法是将少量待确认的细胞注入体内，观察它们重建受损组织的能力。

干细胞很稀少，人体骨髓细胞中只有十万分之一是造血干细胞，而小鼠骨髓细胞中则有万分之一是造血干细胞。

7.2.5 干细胞的衰老

细胞的自然衰老与端粒酶和 DNA 的稳定性相关。

DNA 双链复制时产生的两条新生链中，一条称为前导链，另一条称为后随链。后随链的合成是不连续的，它由冈崎片段连接形成，而且，由于最后一个片段不能启动合成，因此，后随链的长度总是比前导链要短。那么，这种 DNA 链末端短缺部分是否会被填补呢？直线形的染色体两端含有非编码的重复序列，被称为端粒。这种端粒可以通过端粒酶进行复制。端粒酶是一种核蛋白 DNA 聚合酶，可以将端粒延长。端粒酶一旦被表达，就能够使成长中的细胞的染色体端粒长度保持不变。端粒学说认为端粒的长度和端粒酶的激活是细胞获得“永生”的关键因素。在酵母等微生物以及干细胞，还有由干细胞转化而来的肿瘤细胞等衍生物中，都存在端粒酶

活性。但是，定型的干细胞变成单能后代细胞之后，就会丧失端粒酶的活性。因此，正常体细胞并不具有端粒酶活性，每次细胞分裂之后，其染色体上的端粒就会被缩短 50～200bp[⊖]，这种长度变化被称为“分裂钟”（mitotic clock）。端粒的长度约为 9～11kbp，当其长度缩短到 5～7kbp 时，染色体就开始变得不稳定，并会停止复制（见图 7.12）。这就是细胞复制存在 Hayflick 界限的机制，也就是细胞衰老的表现。

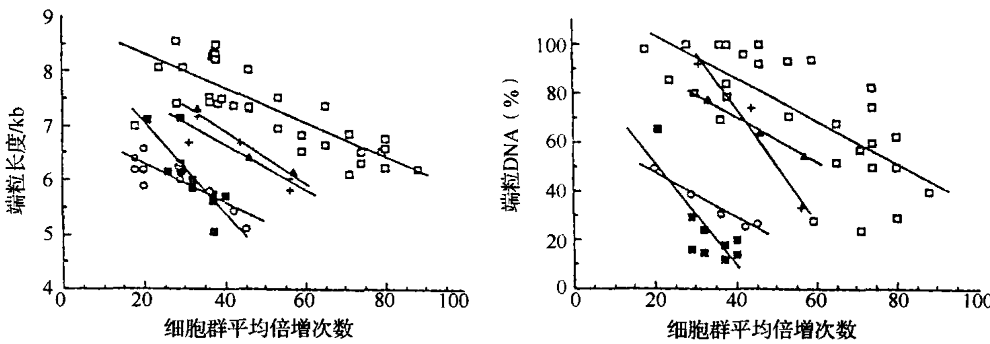


图 7.12 细胞培养过程中随着细胞倍增次数的增加端粒长度缩短（摘自 Harley 等，1990）

造血细胞中存在端粒酶活性，但其活性很低。体外培养实验表明，未成熟造血细胞在发育和细胞倍增过程中其端粒确实有所缩短。干细胞的端粒缩短速度很慢，比体细胞要慢得多。干细胞衰老的数值分析结果表明，成人干细胞存在一定的自我更新能力（见习题 1）。

7.2.6 组织动力学

如图 7.13 所示，人体组织由具有不同发育来源的多种细胞构成，细胞的动态行为及其相互作用决定了整个组织的形态、状态和功能，每个细胞的活性通常都很重要。但是，由于各种细胞活性的变化都需要经过一段时间之后，才能够在组织的功能上体现出来，因此，细胞活性的重要性常被忽视。如下是组织功能动态变化的细胞机制：

1）细胞增殖：通过完整的细胞分裂过程，也就是既有 DNA 合成又有细胞质分裂的过程，实现细胞的复制和增生，从而使细胞数量增加。

2）细胞肥大：只有 DNA 合成却无细胞质分裂，导致细胞体积增大。

3）细胞的成熟和分化：通过基因表达的变化获得某个成熟细胞系的特定功能。

4）细胞凋亡：细胞经历正常的衰老过程，该过程与细胞坏死不同。

5）细胞粘附：细胞与其邻近物质之间的物理连接，该物质可以是相邻的细胞、

⊖ bp 是碱基对。——译者注

细胞外基质或者人造表面。

6）细胞迁移：细胞向特定微环境（即某个部位）的移动。

图 7.14 表示了上述细胞过程。后面会简单介绍目前已了解的有关这些细胞过程的知识，重点放在动态的定量描述上。这些细胞过程构成了生物组织的如下 3 个动态过程：

1）组织发生（Histogenesis）：成熟细胞系来源于组织的干细胞池，组织的整体功能是组织成熟细胞系的所有细胞作用的总和。

2）组织形成（Tissue formation）：发育生物学等领域的研究结果已经阐明了组织的形成过程。随着年龄的变化，各种组织中包含的干细胞、二倍体细胞和多倍体细胞的比例也会发生变化。婴儿等年轻供体的组织中所包含的细胞系中的年轻细胞（即干细胞和二倍体细胞）相对比较多，而老年供体的组织中包含细胞系的晚期细胞（即多倍体细胞）相对比较多。由此可见，体外培养时，取自年轻供体的组织具有较强的扩增潜能，并且能够用于多种不同的细胞治疗。

3）组织修复（Tissue repair）：受损组织的修复过程包括干细胞池的细胞生产、细胞增殖以及细胞分化，进而产生完全成熟的细胞。根据需要，修复过程还可能包括细胞向受损部位的迁移过程。

这些动态过程涉及多种不同类型细胞之间的相互作用。如图 7.14 所示，在几个主要的细胞发育过程中都贯穿了细胞之间的通信与协调。下面分别介绍生物组织的这 3 个动态变化过程的生物学和动力学机制。

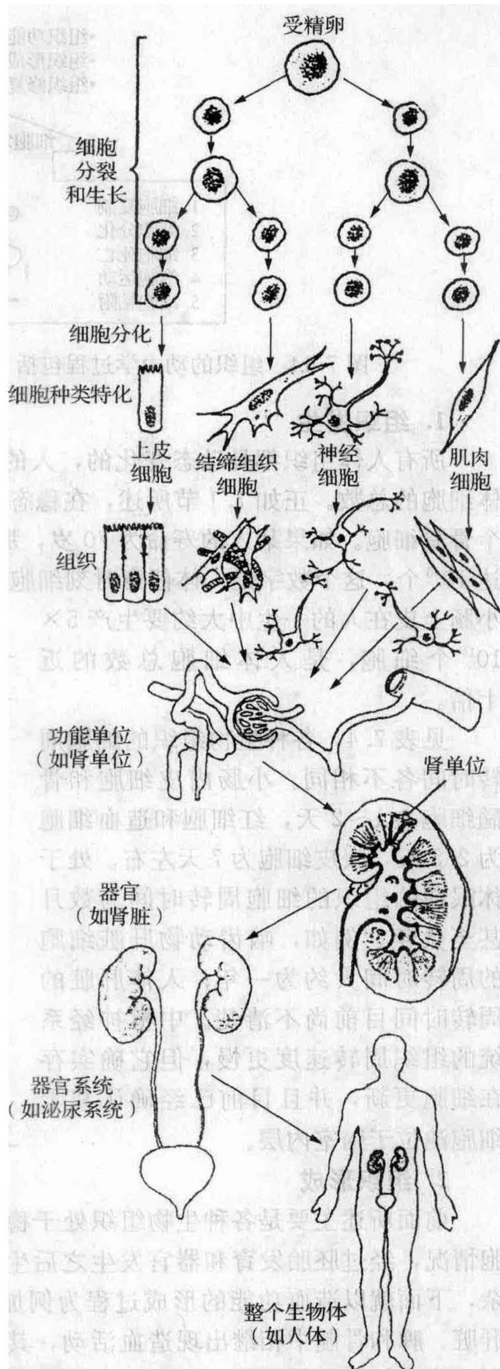


图 7.13 组织的各个细胞层次以及各种细胞的发育来源（摘自 Vander 等，1994）

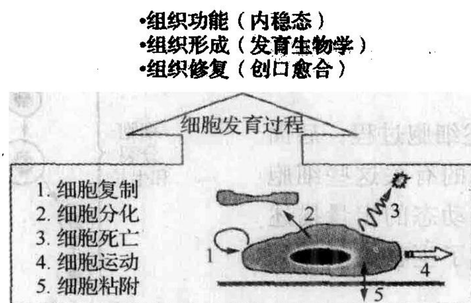


图 7.14 组织的动力学过程包括 3 个动态过程以及各个细胞发育过程

1. 组织发生

所有人体组织都是动态变化的，人的一生中某些器官所产生的细胞数量超过人体细胞的总数。正如 7.1 节所述，在稳态情况下，人体骨髓每天大约要生产 4×10^{11} 个骨髓细胞。如果某人的寿命为 70 岁，那么，其一生中骨髓生产的累计细胞数量可达 10^{16} 个，这个数字是人体任意时刻细胞总数的几百倍！同样，人体第二大增殖组织小肠上皮在人的一生中大约要生产 5×10^{14} 个细胞，是人体细胞总数的近十倍。

见表 7.4，各种生物组织的细胞周转时间各不相同。小肠内皮细胞和骨髓细胞为 1～2 天，红细胞和造血细胞为 2.5 天，表皮细胞为 7 天左右。处于休眠期的组织的细胞周转时间为数月甚至数年。例如，啮齿动物肝脏细胞的周转时间大约为一年；人体肝脏的周转时间目前尚不清楚。中枢神经系统的组织周转速度更慢，但它确实存在细胞更新，并且目前已经确认其干细胞池位于脑室内层。

表 7.4 各种生物组织的细胞更新速度

生物组织	物　种	周转时间/天
红细胞	大鼠	2.5
骨髓细胞	大鼠	1.4
造血细胞	人	2.5
小肠上皮细胞	人	4～6
	大鼠	1～2
表皮细胞	人	7～100
角膜上皮细胞	人	7
淋巴细胞	大鼠（胸腺）	7
	大鼠（脾）	15
上皮细胞	大鼠（阴道）	3.9
	人（子宫颈）	5.7
精原细胞	人	74
肾间质细胞	小鼠	165
肝细胞	大鼠	400～500

2. 组织形成

前面所述主要是各种生物组织处于稳定状态时的干细胞和成熟细胞系的各种细胞情况，经过胚胎发育和器官发生之后生物组织才能逐渐达到稳态，其过程十分复杂，下面就以造血功能的形成过程为例加以说明。脊椎动物发育时，卵黄囊、胎儿肝脏、脾和骨髓中相继出现造血活动，其出现形式在不同脊椎动物物种之间有些差别。造血细胞最早出现在 C4 分裂球的后代 32 细胞胚胎中。当囊胚生长到大约 1000 个细胞时，也就是受精卵经历 10 次倍增之后，囊胚变成球形，然后形成原肠胚，就像用手指捅膨胀的气球一样，内陷部分就是内胚层，它最终将形成消化道。原肠胚

形成之后，外胚层与内胚层之间会形成另一层中胚层，它是由细胞之间的相互作用和可溶性生长因子的活动造成的，它将发育成边缘区。很多定型干细胞都起源于中胚层，包括造血细胞、间充质细胞、肌细胞、肾脏细胞和脊索细胞等。其中造血细胞来自腹侧中胚层。某些造血细胞迁移到卵黄囊形成血岛，血岛细胞大部分是红系细胞（即“原始”造血细胞）。造血细胞的胚胎来源似乎是一种双能细胞，这种细胞既产生管脉系统的内皮组织又产生造血细胞。随着发育的进行，造血干细胞随后出现在胎儿的肝脏。在邻近出生之前，造血干细胞又从肝脏迁移到骨髓，并在胎儿出生后一直位于骨髓中。有趣的是，脐带血中含有造血干细胞，可用于儿童和青少年患者的细胞移植。上述发育过程体现了发育期干细胞分裂的不对称特性及其发育潜能逐渐变窄的过程。此外，发育期干细胞的迁移也非常重要。了解干细胞发育过程的调控机制和动态特性对于组织工程研究至关重要。

3. 组织修复

组织受伤时会诱发创伤愈合反应，其过程包括一系列相互协调的细胞活动。并且，不同年龄阶段的创伤愈合过程有所不同。胎儿期伤口愈合的速度很快且不留疤痕。相反，出生后人体的创伤愈合比较慢，且通常会留下疤痕，虽然有时并不能完全恢复正常组织的结构，但一般都可以获得令人满意的组织修复状态。有些疾病的发生过程也很像创伤愈合过程。例如，各种各样的纤维化疾病就与组织修复过程相似，也会留下疤痕。现在，人们对于干细胞及其后代成熟细胞系的认识越来越深入，对于它们随年龄变化的特性也已经有所了解，这些知识都有助于人类对于创伤愈合现象的进一步认识。下面介绍创伤愈合过程中所发生的一系列事件。

人体受伤之后，血液循环系统中的血小板立即就会粘附到受伤处，并开始控制出血。血小板在数秒时间之内就会被激活，分泌它们所存储的颗粒状物质，并传播出去，同时将更多的血小板聚集到正在形成的血栓处。受伤之后数分钟，伤口周围的血管会收缩，从而进一步减少出血。

创伤愈合的下一步是血小板在伤口处释放因子，促使伤口附近的血管舒张、通透性增加。随后，凝血级联反应被启动，纤维蛋白被凝血酶分解，从而形成纤维蛋白凝块。这种纤维蛋白栓与纤维连接蛋白一起将组织固定住，并形成临时基质。这种基质对于创伤愈合早期炎性细胞的聚集和随后的纤维细胞等附属细胞的迁移，具有重要的作用。

然后，各种炎性细胞迁移到伤口处。嗜中性粒细胞最早从循环血液中转移到伤处。随着嗜中性粒细胞的逐渐脱粒和死亡，巨噬细胞数量不断增加。各种生物组织中都含有巨噬细胞，并且循环血液中的巨噬细胞也会转移到伤处，从而使伤处的巨噬细胞数量增加。它们与嗜中性粒细胞一样，能够吞噬细胞碎片，抵抗微生物的入侵，同时产生趋化因子和促细胞分裂素。这些因子将内皮细胞和纤维细胞引诱到伤处，并刺激它们不断增生。如果伤口处的巨噬细胞受到抑制，那么，创伤的愈合就会受到严重影响。

接着，所谓的肉芽组织形成，它由纤维原细胞、巨噬细胞和旺盛增生的血管系统组成，增生的血管网嵌在纤维连接蛋白、胶原蛋白和透明质酸组成的基质中。纤维细胞开始产生胶原蛋白，大部分是Ⅰ型和Ⅲ型。这些胶原蛋白提高了伤口的抗拉强度。并且，肌纤维细胞此时开始收缩，使伤口缩小，将伤口边缘拉在一起。

在创伤愈合过程中，伤处基质的成分会不断产生变化，包括调控因子的变化和结缔组织蛋白的降解。例如，开始时Ⅲ型胶原蛋白很多，但随着时间的推移它们逐渐被Ⅰ型胶原蛋白取代。这些过程的相互作用决定了伤疤的形态。当表面看起来伤口已经基本愈合时，其实伤处的化学成分和结构仍然在继续发生变化。创伤愈合的最后一步是瘢痕的形成，这个过程可能会持续数个月，其间，经过基质成分的合成和降解作用，创伤处逐渐恢复到正常状态。最后，基质成分和各种细胞的空间分布都基本回复到原有状态。

了解创伤愈合过程对于组织工程的研究工作非常重要，因为体外组织培养时诱导的反应过程与创伤愈合过程很相似。

7.2.7 细胞分化

如图 7.14 所示，细胞发育命运的调节过程决定了组织功能的动态变化，有关调节过程的遗传学、生物化学和动力学上的知识正在不断积累。本章下面将介绍细胞的几个主要发育命运之间相互作用的动力学过程，以及组织工程所面临的挑战。

1. 细胞分化的生物学基础

细胞分化是指细胞经历表型变化之后成为一种特化细胞的过程，这种特化细胞将实现其特有的生理功能。细胞分化起始于细胞的定型，然后是一系列基因表达的调控过程。

“分化”这个术语是指基因表达的差异性，也就是细胞中基因表达的变化。分化使细胞向某种特定功能的状态发展，这种改变通常是不可逆的，它涉及各种基因家族“启动”和“关闭”的精密控制。经过这些基因调控之后，最终的基因表达就与成熟细胞的功能密切相关。动物的细胞分化过程是不可逆的。

例 7.1 请阐述血红细胞的生成过程，产生一个血红细胞需要多长时间？

解： 新生成的血红细胞会取代衰老的成熟血红细胞。人体每天大约需要产生 2000 亿（10^{11}）个血红细胞。如图 7.15 所示，血红细胞生成过程中所经历的各种变化基本已研究清楚。细胞的大小、RNA 和 DNA 的合成速度以及蛋白质的组成都逐步受到调控并发生变化。从原血红细胞（即早期前体细胞）到完全成熟的去核血红细胞，整个分化过程大约需要 180h，也就是约一周时间。其中，在前祖细胞和祖细胞时期，细胞的复制活动最旺盛，但是，一旦到达前体时期，复制活动就立即停止。■

2. 细胞分化的实验测定

利用表面荧光标记和光学显微镜观察细胞形态的方法可以直接考察细胞分化的过程。另外，根据细胞超微结构的变化也可以确定细胞的不同分化时期。

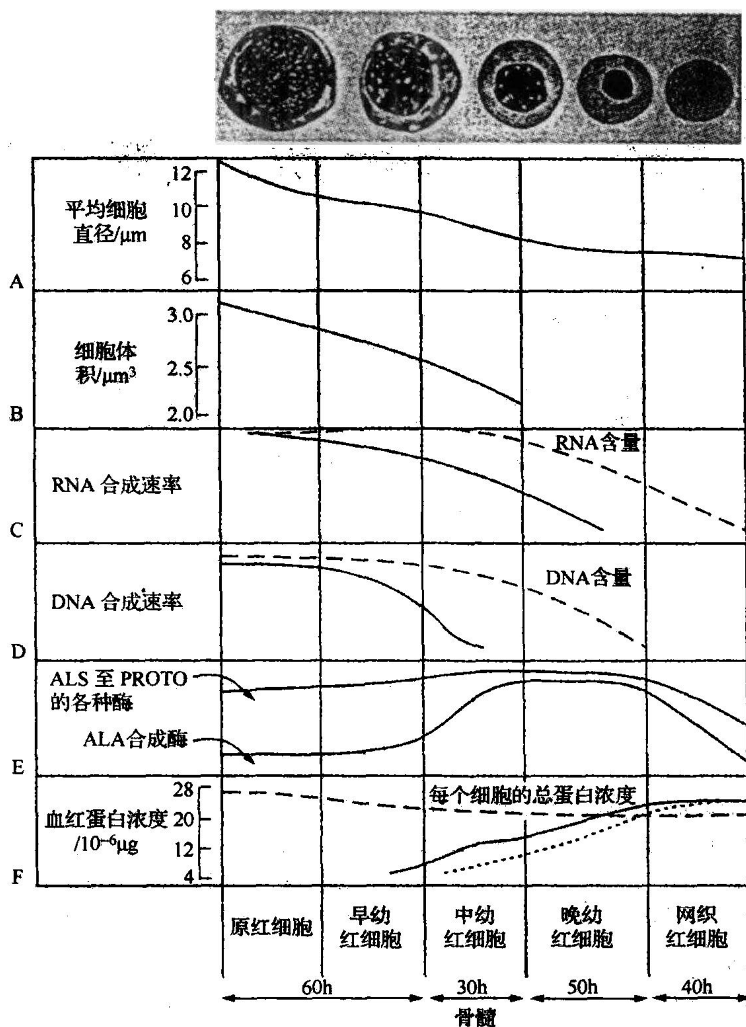


图 7.15 红系血细胞的发育成熟过程（摘自 Granick 和 Levere，1964）

流式细胞仪常用于监测细胞的分化过程。在不同分化阶段，细胞表面的特征蛋白会发生变化，如果将这些表面蛋白作为单克隆抗体荧光标记物的结合位点，就可以利用流式细胞仪跟踪数种表面蛋白的表达情况，从而监测细胞的分化过程。

例如，通过监测转铁蛋白受体 CD71 和血型糖蛋白 A 的表达情况就可以跟踪血红细胞的生成过程。血型糖蛋白 A 是一种红系特异性的表面蛋白，它带负电荷，可以防止高密度血红细胞悬液发生聚集。而转铁蛋白受体在血红蛋白结合铁离子的过程中发挥着重要的作用。图 7.16 显示了血红细胞生成的监测过程。

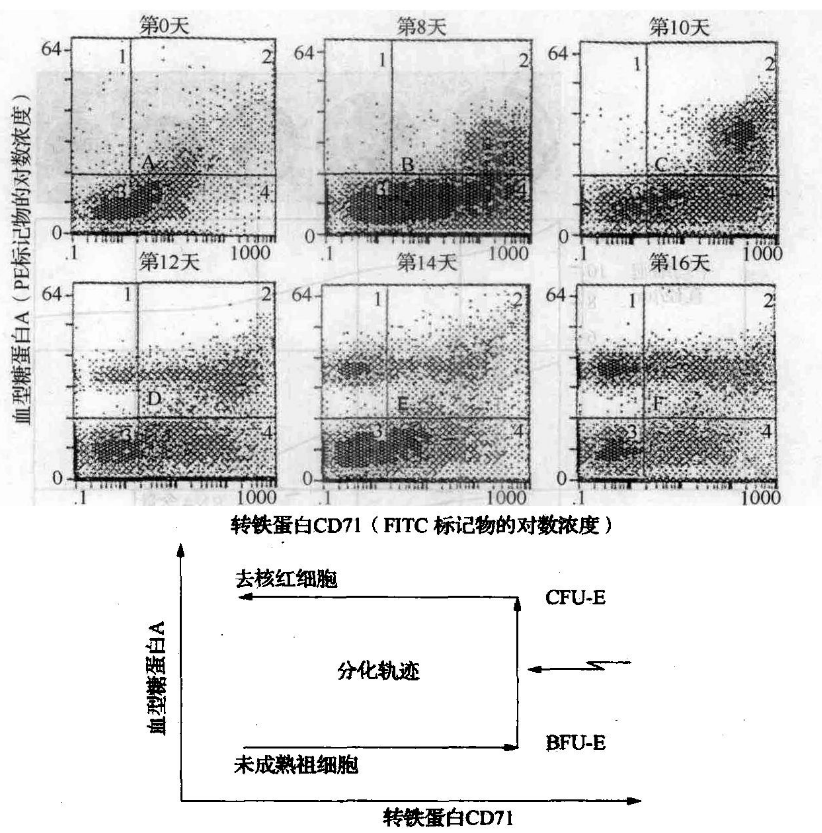


图 7.16　血红细胞的分化过程可以用血型糖蛋白 A 和转铁蛋白 CD71 两个参数来描述

注：血型糖蛋白 A 出现在早期红细胞爆式集落形成单位（Blast Forming Unit-Erythroid，BFU-E），而铁蛋白 CD71 则在 BFU-E 和 CFU-E[⊖] 的祖细胞阶段都有表达。利用流式细胞仪同时测量这两个参数，并绘制双变量散点图，就可以显示出红细胞分化过程中所呈现的 U 形变化轨迹（摘自 Rogers 等，1995）。

3. 细胞分化的动力学模型

细胞分化过程是一个缓慢进行的过程，经常需要几天或者几周才能完成，其复杂的动力学过程可以用如下两种数学方法来描述。

（1）细胞分化的房室模型

描述细胞生长和分化的传统数学方法就是房室模型。分化过程涉及一系列细胞表型和形态的变化过程，这些变化在分化后期越来越明显，它们可以表示为

$$X_0 \rightarrow X_1 \rightarrow X_2 \rightarrow \cdots\cdots X_i \rightarrow \cdots\cdots X_n \rightarrow \text{转分化} \tag{7.1}$$

其中分化阶段数 n 的值可以高达 16～18。

过去的房室模型假设一个分化阶段到下一个阶段的转化代表的是细胞分裂，因此，这种模型将细胞的分化与复制结合在一起，模型的数学方程是如下常微分方

⊖ CFU-E 是红细胞集落形成单位（Colony Forming Unit-Erythroid）。——原文注

程组：

$$\frac{dX_i}{dt}=2k_{i-1}X_{i-1}-k_iX_i \tag{7.2}$$

并且，当前阶段向下一阶段转化的速率正比于当前阶段的现有细胞数量。显然，转化速率是生长因子浓度等变量的函数。

（2）细胞分化的连续模型

另一种建模方法是将分化过程看作连续的过程。细胞一旦定型并进入分化之后，整个分化过程就以恒定的速率进行。这种模型的数学方程是如下一阶偏微分方程组：

$$\frac{\partial X}{\partial t}+\delta\frac{\partial X}{\partial a}=[\mu(a)-\alpha(a)]X \tag{7.3}$$

式中　δ——分化速率；

a——表示细胞分化状态的参数，$a=0$ 表示尚未分化的状态，$a=1$ 表示分化结束的状态；

μ——细胞增生速率；

α——细胞死亡速率。

其中，μ 和 α 取值范围为 0～1，且两者都是 a 的函数。

4. 细胞迁移

（1）细胞迁移的生物学意义

细胞迁移在各种组织正常生理功能和异常病理过程中都具有重要的作用。例如，在器官形成和胚胎发育以及伤口愈合和新生血管形成等组织修复过程中都需要细胞迁移。另外，免疫系统要依靠细胞迁移才能发挥作用，而癌细胞转移等病态过程也是细胞运动产生的结果。因此，细胞迁移是普遍存在的一种细胞活动。

动物细胞在迁移时其形状会发生变化，表面会伸出伪足。细胞伪足有板状伪足和丝状伪足两种，它们都能够动态形成，然后收回。伪足的形成主要是细胞质中局部肌动蛋白聚合的结果，它们由多种蛋白之间相互作用的复杂机制决定。例如，神经细胞突起中的丝状伪足在突起的延伸过程中具有重要的作用，它们有助于微管的装配，而微管是神经细胞的重要组成部分。神经突起中的丝状伪足其实由板状伪足延伸形成，分布于四周，并具有传感器的作用。已有研究结果表明，神经突起的丝状伪足对于神经生长锥的导向起着关键作用。有人发现神经生长锥上的丝状伪足携带了某种细胞粘附分子的受体（参见本章习题 9）。成熟的白细胞也会伸出伪足。最近就有人发现了 T 淋巴细胞中的尾足（uropod）结构，这种尾足在淋巴细胞与上皮细胞相互作用过程中形成，它是 ICAM-1、ICAM-3、CD43 和 CD44 等粘附分子在细胞质中重新分布的结果。T 细胞通过这些尾足与其他 T 细胞产生直接接触，并实现通信。趋化因子等生理因子对于这种尾足的生成具有促进作用。由此可见，细胞质伪足对于不同的细胞具有不同的功能，包括细胞迁移和通信功能。

（2）细胞迁移的动力学模型

无复制细胞群体的整体迁移可以用如下方程描述：

$$\frac{\mathrm{d}X}{\mathrm{d}x}=J \tag{7.4}$$

式中 J——系统边界处的细胞通量矢量，也就是二维系统中单位时间和单位距离上的细胞数量，其单位是“细胞数/距离·时间”；

X——细胞数量；

x——通量的维。

细胞通量与细胞浓度和趋化因子的浓度有关，即

$$J=\text{随机运动}+\text{化学增活运动}+\text{趋化运动}$$
$$=\sigma\frac{\mathrm{d}X}{\mathrm{d}x}-\left(\frac{X}{2}\right)\left(\frac{\mathrm{d}\sigma}{\mathrm{d}a}\right)\left(\frac{\mathrm{d}a}{\mathrm{d}x}\right)+\chi\frac{X\mathrm{d}a}{\mathrm{d}x} \tag{7.5}$$

式中 σ——随机运动系数；

χ——趋化系数；

a——某种趋化因子的浓度。

此式第一项相当于物质传输的扩散项，表示细胞群的随机扩散活动。化学增活项表示细胞运动速度随浓度的变化，一般此项可以忽略不计。如果 $\chi=0$，且化学增活性可以忽略，那么，细胞的运动就是随机游动，类似于分子的自由扩散运动。趋化运动项表示物质浓度梯度引起的细胞运动的偏向，该项相当于物质传输中的对流。

由此可见，随机游动的细胞群可以用类似于描述扩散过程的数学方程来表示。另外，处于迁移状态的细胞的运动特性也可以用单个细胞的运动来表示，如图 7.17 所示，这种运动与上述扩散活动不同。

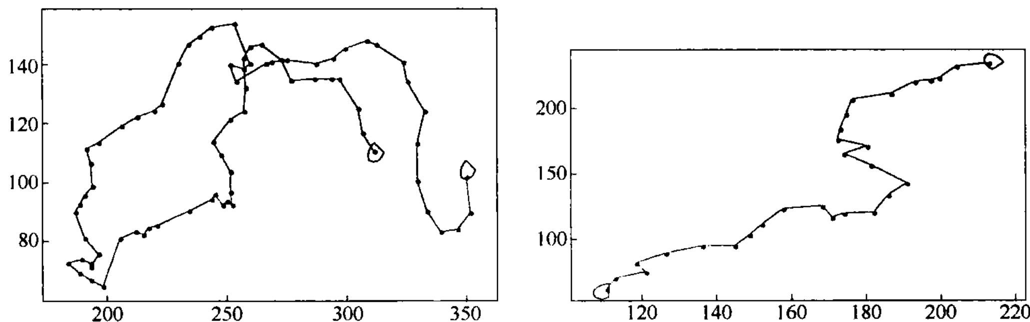


图 7.17 均匀环境中单个嗜中性粒白细胞的随机运动实验记录轨迹
（摘自 Lauffenburger 和 Linderman，1993）

假设细胞的运动速度为 s、保持时间为 p（即细胞保持某个运动方向不变的时间）、并且在趋化因子作用下的方向偏角为 θ；那么，单个细胞的随机运动系数 σ 可以表示为

$$\sigma=s^2p \tag{7.6}$$

表 7.5 列出了这些参数的典型数据。

表 7.5 几种细胞的随机运动速度和保持时间

细胞类型	速度	保持时间
兔类嗜中性粒白细胞	20μm/min	4min
大鼠肺泡巨噬细胞	2μm/min	30min
小鼠成纤维细胞	30μm/h	1h
人体毛细血管内皮细胞	25～30μm/h	4～5h

5. 细胞复制

(1) 细胞周期

如图 7.18 所示，人们对于细胞分裂过程及其分子机制的研究日益深入。细胞分裂周期受到一系列所谓的细胞周期依赖性激酶（cyclin-dependent kinase）的调控，这些激酶存在磷酸化和去磷酸化两种形式，就像双稳态器件一样，可用于组成一系列紧密编排的开关网络，其中的开关或“开”或“关”。当蛋白激酶处于“开”状态时，它们就可以诱导细胞周期所需的生物化学过程。不少综述性文章阐述了哺乳动物细胞周期的各个事件发生顺序。注意，细胞可以从休眠状态（也称为 G_0 期）进入分裂周期，反过来，也可以从分裂周期转变为休眠状态，这些进出细胞周期的转变都由某些重要因素调控。另外，细胞周期的 G_1、S、G_2 和 M 4 个阶段中，S、G_2 和 M 阶段的持续时间比较固定。一旦细胞需要分裂且能够分裂时，它就开始合成 DNA 并随即产生分裂。整个分裂过程具有零阶动力学特性，也就是一旦启动，它就以某个恒定的速度进行。人体细胞最短的周期时间约为 12h，也就是祖细胞的分裂周期。

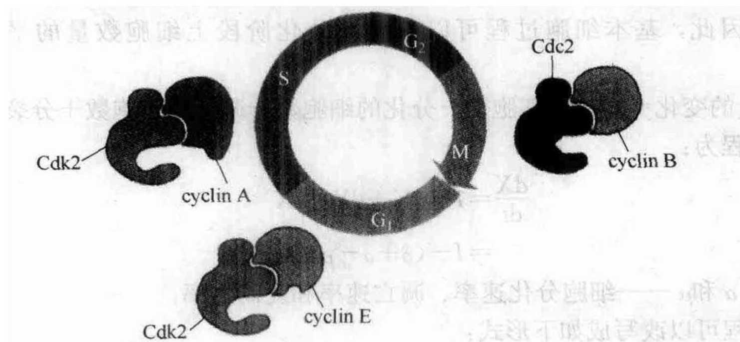


图 7.18 真核细胞的细胞周期（即 G_1、S、G_2 和 M 4 个阶段）及其细胞周期依赖性激酶的示意图

(2) 细胞周期的动力学模型

细胞生长的动力学过程可以用以下几种模型来描述。

1) **指数生长模型**：如果细胞的生长过程不受任何限制，那么，新生细胞的产生速率直接与当前的细胞数量成正比，即

$$\frac{dX}{dt}=\mu X, X(t=0)=X_0 \Rightarrow X(t)=X_0 \exp(\mu t) \tag{7.7}$$

可见，该模型方程的解是一个指数函数。其中生长速率μ等于$\ln(2)/t_d$，t_d是细胞倍增时间。

2）**细胞状态依赖性模型：**如果要描述细胞周期的各个阶段，那么，动态方程中就要结合细胞的状态因素，于是，要用如下时间和细胞状态的一阶偏微分方程：

$$\frac{\partial X}{\partial t}+v\frac{\partial X}{\partial a}=\alpha(a)X \tag{7.8}$$

式中 v——细胞经历细胞周期的速率；

a——描述细胞状态的变量，$a=0$ 表示新生细胞，$a=1$ 表示完成了分裂的细胞；

α——与细胞周期相关的细胞死亡速率。

在适当的初值条件和边界条件下该平衡方程可以求解。

3）**分子机制模型：**细胞周期依赖性激酶的分子级联反应机制已经研究清楚（见第14章），基于这些知识可以建立描述细胞周期的分子机制模型，这种模型含有很多变量，必须利用计算机仿真求解。

6. 细胞发育过程之间的相互作用决定组织的整体动力学特性

细胞分化过程涉及一系列细胞表型和形态的变化，这些变化一般在分化后期越来越明显。其中关键的一步是环境依赖性分化，也就是基因的差异表达，其实就是器官组织形成过程中某种功能性成熟细胞的产生过程。同样，胚胎发育过程也可以描述为一系列基因差异表达的过程。式（7.1）表示了这种过程。

细胞分化过程中的进行性变化一般都与细胞分裂和细胞凋亡等基本“驱动”作用相关。因此，基本细胞过程可以用各个分化阶段上细胞数量的平衡方程来描述，即：

细胞数量的变化＝输入的细胞数－分化的细胞数－凋亡的细胞数＋分裂的细胞增量

其数学方程为：

$$\begin{aligned}\frac{\mathrm{d}X}{\mathrm{d}t}&=I-\delta X-\alpha X+\mu X\\&=I-(\delta+\alpha-\mu)X\end{aligned} \tag{7.9}$$

式中 δ、α和μ——细胞分化速率、凋亡速率和复制速率。

该方程可以改写成如下形式：

$$\frac{\mathrm{d}X}{\mathrm{d}t}+kX=I \tag{7.10}$$

其中$k=\delta+\alpha-\mu$，该参数是时间常数的倒数，表示细胞数量X的动态变化特征。显然，μ/δ和α/μ是决定生物组织中细胞群体产量的两个关键的无量纲比值。

7.2.8 细胞通信

1. 细胞之间的通信方式

如图7.19所示，组织中的细胞相互之间主要通过以下3种方式进行通信：

1）细胞分泌可溶性信号分子，即细胞因子和趋化因子；

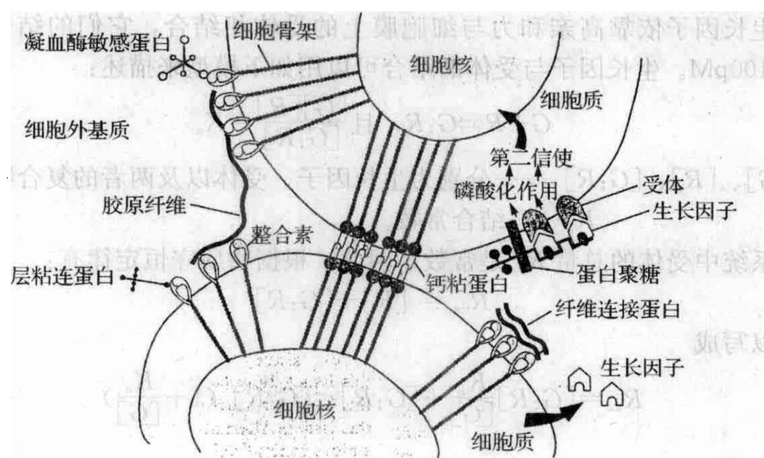


图 7.19　细胞与胞外基质蛋白质之间的相互作用（摘自 Mutsaers 等，1997）

2）细胞之间相互接触，并通过细胞与细胞的直接连接进行通信；

3）细胞利用蛋白质改变微环境的化学性质，即细胞外基质的成分。

这些细胞通信方式的作用时间、作用距离以及特异性都各不相同，因此，每种方式只适合于传递某种特定类型的信息。

2. 可溶性生长因子

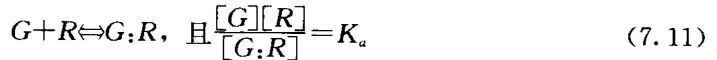
生长因子是分子量为 15～20kD 的小蛋白分子[⊖]。它们的化学性质很稳定，并具有很长的半衰期，除非遭受特异性降解。生长因子最早发现于各种生物体液。例如，促红细胞生成素最初是从尿液中分离获得，而集落刺激因子（colony stimulating factor）则从特殊细胞系的培养液中提取得到。这些蛋白因子可以分离、纯化并定性。随着 DNA 技术的发展，各种生长因子可以作为已知受体的配体直接克隆获得。这种方法已用于分离促血小板生成素（c-mpl 的配体）和干细胞因子（c-kit 的配体）。生长因子都由信号发送细胞生成并分泌，然后传送给靶细胞。

例 7.2　蛋白质的最大分泌速度是多少？

解：通过计算基因转录和翻译的最大速度可以估计细胞中单个基因调控的蛋白质的最大分泌速度。有人用这种方法计算了在高浓度启动子控制下的免疫球蛋白（分子量为 150kD）的生产速度，结果表明，其最大分泌速度为每个细胞每秒分泌 2000～8000 个蛋白分子，相当于每个细胞每小时生产 1pg 蛋白。该计算值与实验测量数据相符。由于生长因子的分子量大约为免疫球蛋白的 1/10，因此，单位时间内每个细胞分泌的生长因子分子数量应该比免疫球蛋白多。不过，由于细胞在任何时候都要同时合成很多种蛋白质，因此，细胞分泌蛋白质的实际速度总是要小于最大速度。■

⊖ 氢原子的质量为 1D。——原文注

各种生长因子依靠高亲和力与细胞膜上的受体相结合，它们的结合常数很小，只有 10～100pM。生长因子与受体的结合可以用如下模型来描述：



$$G+R \Leftrightarrow G:R，且\frac{[G][R]}{[G:R]}=K_a \tag{7.11}$$

式中　$[G]$、$[R]$、$[G:R]$ ——分别为生长因子、受体以及两者的复合物的浓度；

K_a——结合常数。

由于系统中受体的总量 R_{tot} 是常数，因此，根据物质守恒定律有：

$$R_{tot}= [R]+[G:R] \tag{7.12}$$

此等式可以写成

$$R_{tot}=[G:R]\frac{K_a}{[G]}+[G:R]=[G:R]\ (1+\frac{K_a}{[G]}) \tag{7.13}$$

即

$$\left(\frac{[G:R]}{R_{tot}}\right)=\frac{[G]}{([G]+K_a)} \tag{7.14}$$

如图 7.20 所示，要使刺激信号发生作用，受体的占有率 $[G:R]/R_{tot}$ 必须达到 0.25～0.5 的水平，因此，即使 10pM 这样小的生长因子浓度也足以诱发细胞的响应。

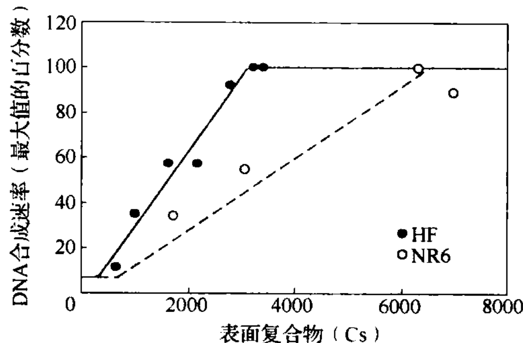


图 7.20　细胞表面表皮生长因子复合物稳态浓度与细胞的 DNA 合成响应之间的关系

注：其中，空心圆点表示鼠成纤维细胞（即 NR6），实心圆点表示人成纤维细胞（即 HF）

在很多情况下，受体与配体的复合物会被内化，其内化的时间常数一般是 15～30min。有些生长因子摄入速率的绝对值已被测定。例如，未成熟的造血细胞以每天每百万个细胞 10～100ng 的速度消耗白细胞介素 3（即 IL-3）和干细胞因子 SCF。此外，实验结果表明，在复合的细胞培养基中，消耗 10000～70000 个生长因子分子才能够刺激细胞分裂。

例 7.3　可溶性信号分子的作用距离有多远，其传播需要多长时间？

解： 利用球形物体分泌物质的简单扩散模型就可以计算细胞信号分子可以传播的最远距离。稳态时，距离细胞 r 处的分泌物的浓度 c 是距离 r 的函数，即

$$\frac{c}{K_a}=\alpha\frac{R}{r}，其中\ \alpha=\frac{(R^2/D)}{(K_aR/F)} \tag{7.15}$$

式中　R——细胞半径；

F——分泌速率；

D——生长因子的扩散系数。

于是，$c=K_a$处的临界距离 r 就是

$$r_{临界}/R=\alpha \tag{7.16}$$

其中，最大分泌速率 α 等于两个时间常数之比，即扩散时间常数（R^2/D）和分泌时间常数（K_aR/F）。由于理论上需要无限长时间才能达到稳态，因此，比较合理的一种方法是计算信号分子浓度达到其最终稳态值的一半时的传播距离。这样计算得到的信号传播时间常数约为 20min，最大距离约为 200μm。这个距离与分泌速率 F 成正比，而与亲和力 K_a 则成反比。■

如图 7.21 所示，生长因子与其受体结合之后，就会触发复杂的信号转导过程。一般，受体复合物会发生变化，并由此诱导细胞内的一系列酶促反应，很多情况下，这些反应与 Janus 激酶（即 JAK）相关。这些激酶通过信号传导和转录激活因子的作用，将信号传送到细胞核内。其中的动力学过程很复杂，表皮生长因子的详细信号传导过程已基本弄清楚。

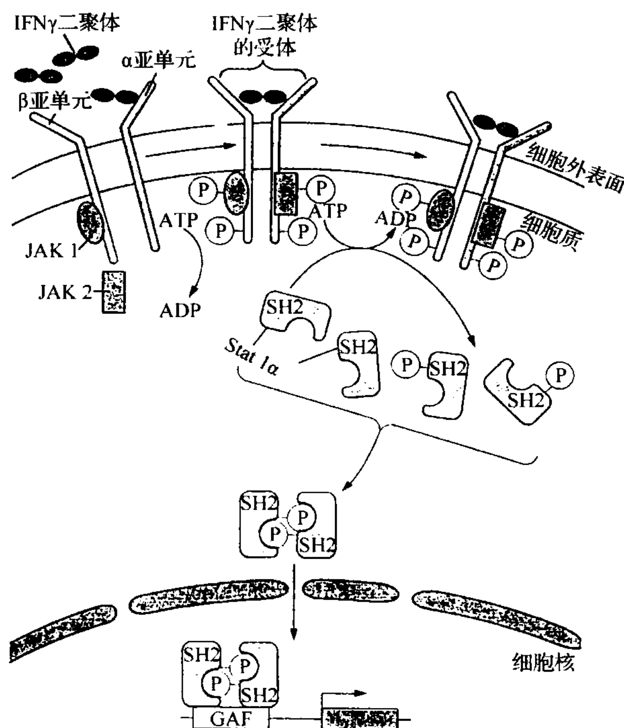


图 7.21　干扰素-γ 的信号转导途径示意图（摘自 Lodish 等，1995）

3. 细胞之间的直接接触——不可溶因子之一

细胞表面存在一类被称为细胞粘附分子（CAM）的蛋白质，其中包括构成粘合带和桥粒的钙粘蛋白以及构成缝隙连接的连接蛋白等。这些蛋白分子形成了细胞与细胞之间的直接接触。

这类分子中的某些分子被称为细胞连接分子，它们在相邻细胞之间形成连接，允许细胞质之间的直接通信。这种连接构成的通道的直径一般约为 1.5nm，可以允许分子量小于 1kD 的分子通过。

最近有越来越多的文章阐述流体力学剪切力对细胞和组织功能的影响。在细胞中，细胞骨架具有重要的力学作用。生长因子与受体结合时会引起细胞的拉伸，从而通过细胞骨架将信号传送到细胞核。整合素就是各种重要信号的“力学传感器”。此外，根据各个细胞在组织中所处的不同位置，通过微弱的力学作用，细胞之间还可以进行长距离通信。因此，细胞外微环境的力学特性与细胞本身的力学特性一样重要。

4. 细胞与基质之间的相互作用——不可溶因子之二

细胞外基质 ECM 就像编织物、胶水、支撑物和凝胶组成的复合体，它将组织中的各种细胞及其构架连接在一起。ECM 具有很多功能，它为组织提供力学支撑，为细胞提供迁移所依附的基底，还为细胞之间的通信提供信号分子传导的场所。因此，ECM 既具有结构功能又具有信息功能。而且，它是动态变化的，会不断发生改变。例如，ECM 的成分会被金属蛋白酶降解；心肌中大约 3%的基质每天要被更新。

ECM 中含有各种成分，它们的结构和调控功能各不相同。细胞表面有许多可以促进细胞与基质之间通信的粘附分子和 ECM 受体分子。这些信号分子包含了细胞迁移、复制、分化和凋亡的指令信息，它们的特性受到 ECM 成分的调控，而 ECM 成分反过来又受到各种细胞的调控。ECM 引领着各种细胞的发育命运，并且是细胞之间通信的工具。ECM 中的信号分子比较稳定，与那些通过扩散传递信息的生长因子相比，它们的特异性更强，信号传导能力也较强。

表 7.6 归纳了 ECM 所包含的各种组成成分及其功能。

表 7.6 细胞外基质的各种组成成分[①]

组 分	功 能	分 布
胶原蛋白	组织构架、拉伸强度、细胞与基质的相互作用、基质之间的相互作用	无处不在
弹性蛋白	组织构架和特性	具有弹性的组织，如肺、血管、心脏、皮肤等
蛋白聚糖	细胞与基质的相互作用、基质之间的相互作用、细胞增殖、各种生长因子的连接与存储	无处不在
透明质酸	细胞与基质的相互作用、基质之间的相互作用、细胞增殖、细胞迁移	无处不在
层粘连蛋白	基底膜的组成成分、细胞迁移	基底膜
表皮整联配体蛋白	上皮细胞基底膜的组成成分	基底膜
副层连蛋白	基底膜的组成成分	基底膜

（续）

组　　分	功　　能	分　　布
纤维连接蛋白	组织构架、细胞与基质的相互作用、基质之间的相互作用、细胞增殖、细胞迁移、调理素	无处不在
玻璃连接蛋白	细胞与基质的相互作用、基质之间的相互作用、止血	血液、伤口形成处
纤维蛋白原	细胞增殖、细胞迁移、止血	血液、伤口形成处
微纤维蛋白	弹性纤维的微纤维成分	具有弹性的组织，如肺、血管、心脏、皮肤等
固生蛋白	调控细胞与基质的相互作用	基质重塑过程中瞬时表达
骨连蛋白 SPARC②	调控细胞与基质的相互作用、抗粘连、抗增殖	基质重塑过程中瞬时表达
凝血酶敏感蛋白	调控细胞与基质的相互作用	血小板 α 颗粒
各种粘附分子	调控细胞与基质或者细胞与细胞之间粘附能力的细胞表面蛋白、跨膜信号分子的介质	无处不在
假性血友病因子（即 vWF）	调控血小板的粘附性、凝血因子Ⅷ的载体	血浆蛋白、内皮下膜

①摘自 Mutsaers 等，1997。
②SPARC 是富含半胱氨酸的酸性分泌蛋白。

例 7.4　不同细胞功能的实现需要多少受体位点？

解：将精氨酸—甘氨酸—天门冬氨酸（RGD）三肽序列以不同的密度固定在细胞表面，就可以检测纤维细胞表面 RGD 结合位点密度对于细胞粘附、伸展和生长的影响。实验结果表明，平均间隔为 440nm 的受体分布密度对于细胞粘附和伸展已经足够，160nm 的分布间隔则可以实现局部点上的粘附形成。

如今，组织工程的很多研究工作致力于人工构建细胞外基质。人造 ECM 一般用聚合物材料制作，有些材料是生物可降解的，在机体生长过程中人工 ECM 可以被逐渐取代，从而重建组织功能。不过，由于天然 ECM 的大部分功能目前尚不清楚，而且它承担着细胞之间的双向通信功能，因此，人工 ECM 很难设计。如表 7.7 所示，只有细胞自身能够提供完整的有关 ECM 功能的信息。■

表 7.7　调控细胞与细胞外基质之间相互作用的各种粘附分子

粘 附 分 子	配　　体
整合素	
$\alpha_1\beta_1$	Ⅰ、Ⅳ、Ⅵ 型胶原蛋白，层粘连蛋白
$\alpha_2\beta_1$	Ⅰ～Ⅳ、Ⅵ 型胶原蛋白，层粘连蛋白
$\alpha_3\beta_1$	Ⅰ型胶原蛋白，层粘连蛋白、纤维连接蛋白、副层连蛋白、表皮整联配体蛋白
$\alpha_4\beta_1$	纤维连接蛋白$_{ALT}$、VCAM-1、凝血酶敏感蛋白
$\alpha_5\beta_1$	纤维连接蛋白、凝血酶敏感蛋白
$\alpha_6\beta_1$	层粘连蛋白
$\alpha_v\beta_1$	纤维连接蛋白
$\alpha_L\beta_2$	ICAM-1、ICAM-2、ICAM-3
$\alpha_M\beta_2$	ICAM-1、iC3b、纤维蛋白原、X 因子、变性蛋白

（续）

粘附分子	配　　体
$\alpha_x\beta_2$	纤维蛋白原、iC3b、变性蛋白
$\alpha v\beta_3$	玻璃连接蛋白、纤维蛋白原、纤维连接蛋白、凝血酶敏感蛋白
$\alpha v\beta_5$	玻璃连接蛋白
$\alpha v\beta_6$	纤维连接蛋白
$\alpha_0\beta_4$	层粘连蛋白
$\alpha_4\beta_7$	纤维连接蛋白$_{ALT}$、VCAM-1、MAdCAM-1
$\alpha_{1th}\beta_3$	纤维蛋白原、纤维连接蛋白、玻璃连接蛋白、vWF
LRI^h	纤维蛋白原、纤维连接蛋白、玻璃连接蛋白、vWF、IV型胶原蛋白、副层连蛋白

7.3 物理学问题

7.3.1 组织功能亚单位的形成

如表7.8所示，人体一共有11个主要器官系统，其中肌肉与骨骼一般合在一起，称为肌肉骨骼系统。这些器官系统实现呼吸、消化和机械运动等生理机能，每个系统都由多个器官组成，图7.22显示了消化系统所包含的主要器官。各个系统及其器官都具有自我平衡功能，这些功能由生理需求决定，就像商品的“技术指标”一样，表7.9举例说明了男性人体的生理学“技术指标”数据。

表7.8　人体的主要器官系统

系统名称	组成部分	功　　能
循环系统	心脏、血管、血液，有时还包括淋巴管和淋巴	向全身组织输送血液
呼吸系统	鼻、咽、喉、气管、支气管、肺	交换二氧化碳和氧气，调节氢离子浓度
消化系统	口、咽、食道、胃、肠、唾液腺、胰腺、肝脏、胆囊	有机营养成分、盐和水的消化与吸收
泌尿系统	肾脏、输尿管、膀胱、尿道	通过盐、水和有机废物的排泄来调控血浆成分
肌肉骨骼系统	软骨、骨骼、韧带、肌腱、关节、骨骼肌	起支撑和保护作用，并实现人体运动和血细胞的生产
免疫系统	脾、胸腺等淋巴组织	抵制异物的侵入，将细胞外液返回血液，生成白细胞
神经系统	脑、脊髓、中枢神经与外周神经、感觉器官	调控机体的各种活动，检测机体内部环境和外部环境的各种变化，实现意识、学习和认知过程
内分泌系统	包括所有分泌激素的腺体：胰腺、睾丸、卵巢、下丘脑、肾脏、垂体、甲状腺、甲状旁腺、肾上腺、肠、心脏、胸腺、松果腺体等	调控机体的各种活动

（续）

系统名称	组成部分	功　能
生殖系统	男性包括睾丸、阴茎、输精管和附属腺；女性包括卵巢、输卵管、子宫、阴道、乳腺	男性：产生并输送精子；女性：产生卵子、给胚胎和胎儿的发育生长提供营养环境；给胎儿提供营养
皮肤系统	皮肤	保护机体免受损害和脱水，抵制异物的侵入，调控体温



图 7.22　消化系统的结构与功能（摘自 Vander，1994）

表 7.9 标准美国男性人体的生理学数据①

年龄	30 岁
身高	5ft8in，即 1.86m
体重	150lb，即 68kg
体表面积	19.5ft^2，即 1.8m^2
正常体温	37.0℃
正常皮肤平均温度	34℃
热容	0.86cal/g/℃
一些容量数据	
机体脂肪含量	10.2kg，或者体重的 15%
皮下脂肪层厚度	5mm
体液	约为 51L，或者体重的 75%
血容量	5.0L，其中包括血液的有形成分、各种血细胞、血浆。红细胞占全血容积的百分比约为 0.43
肺	
总肺容量	6.0L
肺活量	4.2L
潮气量	500mL
死腔	150mL
物质传输面积	90m^2
静息时的物质平衡与能量平衡	
能量转化速率	72kcal/h，即 1730kcal/天（即 40kcal/m^2/h）
氧气消耗速率	250mL/min（呼吸商为 0.8）
二氧化碳产生速率	200mL/min
心率	65 次/min
心输出量	5.0L/min
人体血压	120/80mmHg

① 摘自 Lightfoot，1974。

通过分析人体生理学数据大致的数量级，我们可以获得许多重要的信息，这是正确预测未来一段时间内人体组织功能状况的基本方法。不过，分析计算并不能代替细致的实验研究。

正常的器官有能力响应某些变化因素的作用，可以运行于偏离基本数据的某些状态。例如，在剧烈运动时，血液循环的速率和血液在各种器官中的分布都与静息状态有明显差别。同理，在异常的造血需求作用下，如出现感染或者镰刀型血红细胞贫血症时，人体的血细胞生成速率会显著超过图 7.3 所示的正常水平。

各种器官都由功能亚单位组成，如肺中的肺泡和肾脏的肾单位（见图 7.23）都是功能亚单位。这些亚单位由不同种类的细胞一起构建而成，从而实现特定的组织功能。如果将功能亚单位中的每一种细胞分离出来，那么，组织特异性功能就会随之丧失，不过，这种分离提纯的细胞可用于研究特定细胞的性能。

总而言之，组织的功能是细胞之间相互作用的结果；功能亚单位的大小约为 100μm 数量级；而细胞的大小为 10μm 数量级；每个器官由数千万至数亿的功能亚单位组成。人体各种器官的大小是长期进化的结果，它给组织工程人工器官的制造提

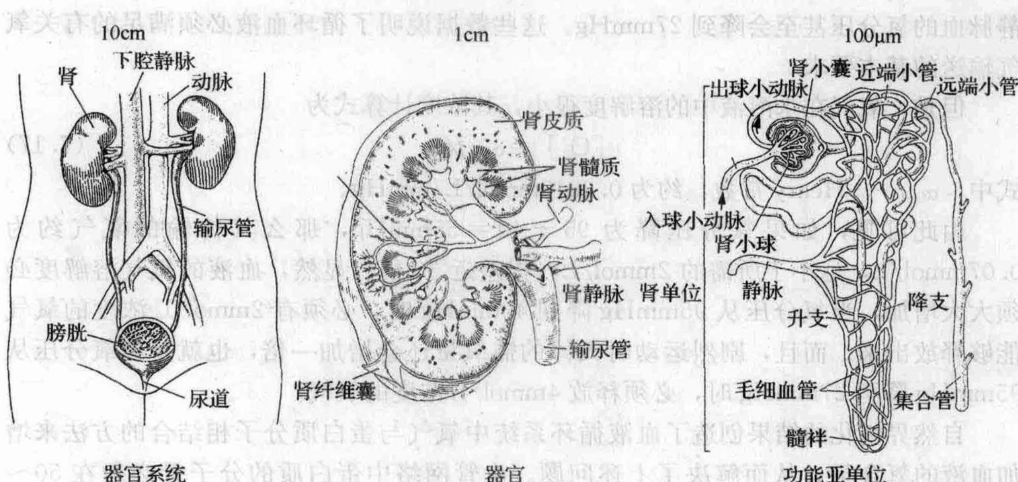


图 7.23　器官系统、器官与功能亚单位

出了难题，体外重建的功能性组织必须具有相应的尺寸。

器官组织的微环境非常复杂。为了实现组织的体外重建，必须精确复制其微环境的动力学特性、化学特性以及几何结构特性，这很困难。本章下面将介绍描述微环境的定量方法，利用这些方法可以更好地了解微环境中的关键问题，获得解决问题的方法，并用于指导实验研究。

组织的所有微环境通过微循环与人体的总环境相连。除了极少数例外情况，人体中所有具有代谢活性的细胞基本上都分布在毛细血管周围数百微米距离之内。毛细血管提供的灌流功能将每个细胞与氧气供应源、二氧化碳吸收源（即肺脏）、营养供应源（即小肠）以及废物清除器（即肾脏）等器官相连。生物反应器的开发就是要在体外实现这些功能。这种培养装置必须正确模拟人体的呼吸功能、肠胃功能和肾功能，而且，还必须满足微环境的条件，也就是在大于 100μm 的尺度上提供均匀的灌流特性，这些都是苛刻的设计要求。

7.3.2　根据人体生理学数据计算组织功能的参数

组织工程中大部分实用的分析结果都是根据生理学和细胞生物学的数据（见表 7.9 的人体数据等）进行近似计算获得，这些计算对于解释器官的生理功能并指导实验过程非常有用。下面举例说明分析计算的方法。

1. 血液的呼吸功能

有关血液的生理呼吸功能已有详细的计算数据，根据人体组织的生理学数据可以直接分析组织的基本功能特性和生物设计中所面临的问题。

人体血液每分钟大约需要输送 10mmol 氧气，人体血流的总循环速率大约为每分钟 5L，因此，血液传送氧气的浓度约需 2mmol/L。刚离开肺部的血液中的氧分压约为 90～100mmHg。静息状态下，静脉血的氧分压只有 35～40mmHg；剧烈运动时，

静脉血的氧分压甚至会降到 27mmHg。这些数据说明了循环血液必须满足的有关氧气输送的基本要求。

但是，氧气在水溶液中的溶解度很小，其浓度计算式为

$$[O_2] = \alpha_{O_2} pO_2 \tag{7.17}$$

式中 α_{O_2}——Henry 常数，约为 0.0013mmol/L/mmHg。

由此可见，如果氧分压降为 95 − 40 = 55mmHg，那么其传输的氧气约为 0.07mmol/L，远小于所需的 2mmol/L，相差近 30 倍！显然，血液的氧气溶解度必须大大增加，当氧分压从 95mmHg 降到 40mmHg 时，必须有 2mmol/L 浓度的氧气能够释放出来。而且，剧烈运动时人体的需氧量还会增加一倍，也就是当氧分压从 95mmHg 降至 27mmHg 时，必须释放 4mmol/L 浓度的氧气。

自然界进化的结果创造了血液循环系统中氧气与蛋白质分子相结合的方法来增加血液的氧含量，从而解决了上述问题。血管网络中蛋白质的分子量必须在 50～100kD 范围内。如果每个蛋白质分子上只有一个氧分子的结合位点，那么，要达到 10mmol/L 的氧浓度所需的蛋白质浓度就是 500～1000g/L。从渗透压上看，这个浓度太高，而且，此时血液的粘性会达到实际循环系统血液的十倍，这显然不可行。另外，循环系统中的蛋白酶会缩短含有这种蛋白质浓度的血浆的半衰期。

如果每个携带氧气的蛋白质分子有 4 个结合位点，那么，蛋白质的浓度就可以降至 2.3mmol/L，再将蛋白质置于红细胞内，这样就解决了粘性和蛋白质降解这两个问题。这就是血红蛋白的主要特性。如图 7.24 所示，血红蛋白氧合特性的动力学模型研究结果表明，其氧合能力可以满足静息状态和剧烈活动状态的人体氧气输送需求。

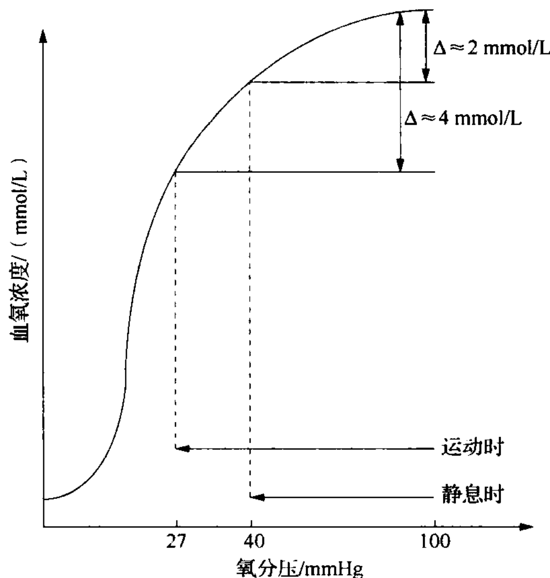


图 7.24 血红蛋白与氧分子结合的曲线以及氧气传输的示意图

注：血液流过组织时产生的氧浓度变化（即氧气传输量）是血液氧分压下降量的函数。

2. 人体骨髓培养的灌流速率

在设定体外培养细胞的条件时，培养基更新速率的确定至关重要。一般，刚开始时，实验设计最好参照体内生长环境的数据，因此，我们首先必须了解细胞在体内生长的动力学特性。

在人体内，每毫升骨髓的血液灌注速率约为 0.08mL/min，而每毫升骨髓约含 5 亿个细胞，因此，按照细胞数量计算，相当于每天每千万个骨髓细胞的血液灌注速率为 2.3mL（即 2.3mL/10^7细胞/天）。20 世纪 70 年代中后期，鼠类骨髓单核细胞的培养就已开发成功，并且这种细胞培养的存活期很长。但是，20 世纪 80 年代初，利用同样的培养方法进行人类骨髓细胞培养时却不成功，当时采用的方法是每周更换一次培养基。

那么，这种体外培养的培养基更新速率与体内血液灌注速率是否相当呢？我们先将体外培养基的更换速率换算成与上述体内计算相同的量，再比较两者的差别。通常，体外培养的初始细胞密度约为 10^6 个/mL，因此，10mL 约包含 10^7 个细胞，而 10mL 培养液中含 20%（vol/vol）血清培养基；由此可见，每天更换一次培养基才相当于每天每千万个细胞 2mL 的灌流速度，才能够与上述体内计算结果相当。

于是，在 20 世纪 80 年代末，根据这种灌流速率和细胞密度进行的细胞培养实验就取得了成功，各种人体骨髓细胞的培养从此迅速发展。经过培养规模的调整，这些培养实验可以生产适合临床应用的细胞数量，如今相关临床试验正在进行之中。由此可见，体内与体外组织之间动力学相似性的分析可以指导具有临床意义的细胞培养方法的开发，这些分析的依据就是人体组织的生理学数据。

上述血红蛋白氧合特性和骨髓细胞培养灌注特性这两个例子说明了近似计算在指导组织工程的初期研究工作中具有重要作用，其基础是全面正确的人体生理学数据。根据这些数据，可以计算时间常数、距离常数、流量、速率、浓度等参数，这些特征参数的数量级是组织工程研究工作的重要依据。

7.3.3 三维结构中的物质传输

各种生物系统中物质传输的计算是组织工程的基础，设计和使用细胞治疗产品（或体外器官）时都必须知道物质的传输速率。人体生理学数据只提供了整体的物质传输信息，而组织工程需要生物组织的最小生理单位的毛细血管床的详细物质传输数据。物质传输包括进出组织的营养物质和废物的扩散和对流以及组织的营养物质消耗和废物的产生等。如表 7.10 所示，血管系统的血液对流速度变化极大，主动脉中通常为 140cm/s，而到了毛细血管床却只有 10^{-3} cm/s。对流由血压差产生，它是血管系统中物质传输的主要方式；而扩散是由浓度梯度产生，它是组织中物质传输的主要方式。扩散可以用 Fick 定律来描述，经过适当的修正，Fick 方程可以描述如图 7.25 所示的矩形、圆柱形和球形 3 种不同结构物体中的物质传输现象。

表 7.10　血管系统中血液的最大对流速度
（数据取自 DeWitt 和 Wechsler（1988））

血管类型	流速/(cm/s)
主动脉	140±40
颈总动脉	100±20
椎动脉①	36±9
表浅股动脉	90±13
肝脏血窦②	$10^{-2}\sim10^{-3}$

① 数据取自 Jager 等（1985）。
② 数据取自 McCuskey（1984）。

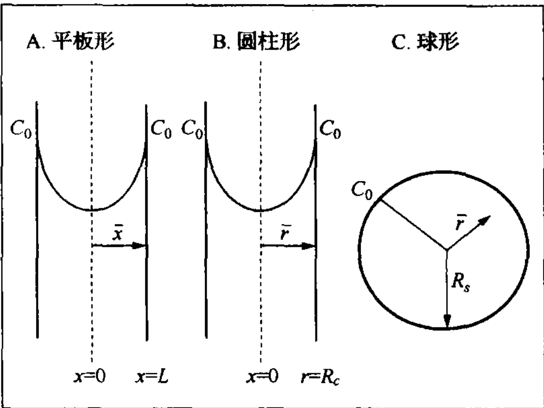


图 7.25　3 种不同形状物体的物质传输方程
（根据 Macdonald 等（1999）绘制）

用于干细胞体外扩增的生物反应器设计的关键就在于提供有利于分化的局部环境的同时，要保持足够的物质传输速率。组织或细胞团中氧气和葡萄糖等小分子营养物质的传输机制主要是扩散。扩散传输的距离（例如从血流中扩散出来的氧气在被组织细胞完全消耗之前所能穿透的距离）取决于物流中所含物质的容积浓度 C_0、细胞消耗该物质的速率 Q_i、物质在组织中的扩散系数 D_i 以及系统的几何形状。细胞消耗物质的速率 Q_i 通常是细胞团中该物质浓度 C 的函数。最常用的物质消耗速率方程就是 Michaelis—Menton 动力学方程，即

$$Q_i=(V_{\max}\times C)/(K_m+C)$$

当物质浓度较高时，即 $C\gg K_m$，该方程可以简化为零阶动力学方程；当物质浓度较低时，即 $C\ll K_m$，该方程变为一阶动力学方程。在正常生理状态下，许多重要的营养物质都符合 $C\gg K_m$ 的条件，因此，可以使用零阶方程。

营养物质在三维细胞团中的扩散距离可以用数学模型来计算。稳态扩散与物质消耗之间的平衡方程为

$$D_i\nabla^2 C=Q_i \tag{7.18}$$

式中　C——细胞团中营养物质的浓度。

图 7.25 显示了平板形、圆柱形和球形 3 种简单形状细胞团孵育在浓度为 C_0 的培养基物质中时的扩散情况，这 3 种几何形状的扩散方程都有解析解。将式（7.18）的梯度算子分别根据 3 种形状展开，可得如下方程：

$$平板形：D_i\frac{d^2C}{dx^2}=Q_i \tag{7.19a}$$

$$圆柱形：D_i\frac{1}{r}\frac{d}{dr}\left(\frac{rdC}{dr}\right)=Q_i \tag{7.19b}$$

$$球形：D_i\frac{1}{r^2}\frac{d}{dr}\left(\frac{r^2dC}{dr}\right)=Q_i \tag{7.19c}$$

图 7.25 分别显示了各种形状下的距离变量的定义。求解这些物质传输微分方程的标准方法是将各个变量转化为取值范围为 0～1 的无量纲变量。这样，方程中的各个量就都成为相对值，很容易计算，所得到的解函数具有普遍适用性，可以用一组曲线图来表示。对于每一种几何形状，浓度 C 按照其最大值 C_0 进行归一化变换，即 $\overline{C}=C/C_0$。距离则通过除以扩散路径的长度进行归一化，对于平板形是 $\overline{x}=x/L$；对于圆柱形是 $\overline{r}=r/R_c$；对于球形是 $\overline{r}=r/R_s$。这样定义之后，再加上边界条件就可以求解方程。这 3 种形状方程的边界条件是：中心处的物质流量为 0，即在 $\overline{x}=0$ 和 $\overline{r}=0$ 处有 $d\overline{C}/d\overline{x}=0$，$d\overline{C}/d\overline{r}=0$；并且在细胞团的表面，物质浓度为最大值，即在 $\overline{x}=1$ 和 $\overline{r}=1$ 处有 $\overline{C}=1$。

经过变量的无量纲处理之后，3 种形状方程的解可以用同一种通用形式来表示，即

$$\overline{C}=1-\frac{\phi^2}{2}(1-\overline{x}^2) \tag{7.20}$$

所有的系统参数都集中在一起，并用无量纲参数 ϕ^2 表示，ϕ^2 被称为 Thiele 模数。该模数表示化学反应或者物质扩散的相对速率，对于上述 3 种形状其定义略有不同，即：

$$\phi^2_{平板}=\frac{QL^2}{C_0D},\quad \phi^2_{圆柱}=\frac{QR_c^2}{2C_0D},\quad \phi^2_{球形}=\frac{QR_s^2}{3C_0D}$$

7.3.4 组织的微环境——细胞治疗和生物反应器设计的原则

以上介绍了组织功能的整体情况，这些功能都建立在细胞发育过程的相互作用之上。并且，还讲述了定量描述这些细胞发育过程的方法，细胞发育过程由复杂的基因周期控制。

下面将介绍物质的合成。组织工程的另一项工作就是要弄清楚生物组织的功能是如何建立、重建和改变的。首先了解一些基本概念：

1）组织的发育程序和创伤愈合响应都需要 DNA 遗传信息的系统性展开与调控，这些 DNA 信息的展开通过许多遗传“子程序”和“程序”的执行来完成。参与发育和创伤愈合的细胞还需要邻近区域其他细胞等物质的详细活动信息，因此，必须了解细胞之间的通信和干细胞微环境等概念。

2）器官形成或创伤愈合之后，器官的整体功能与多种类型细胞的功能调控密切相关。生物组织的功能由很多细胞集群共同完成，这些细胞集群被称为组织的功能亚单位。

3）微环境是每个细胞功能的主要影响因素，微环境的作用距离约为 100μm。

4）微环境包含了邻近区域的细胞、化学环境和局部几何形状等特征。其中，邻近的细胞会产生细胞之间的连接、可溶性生长因子等物质；化学环境是指细胞外基质及其营养环境的动态变化。

1. 体内细胞功能——细胞的组织微环境以及与其他器官的通信

组织功能成功实现的关键是必须具有符合生理需求的细胞生长环境，使细胞能够表达所需的组织功能。虽然并非惟一的方法，但是一般体外培养环境都模拟体内的生理环境。那么，如何认识并重建符合生理要求的组织微环境呢？

如图 7.2 所示，每个细胞与其相邻环境以及其他组织之间的通信可以用人体拓扑结构的示意图来描述。该图的中心是 DNA，它包含了组织工程设计人员希望表达和操纵的遗传信息。细胞所处的微环境具有重要的时空特性。图 7.26 显示了 3 种组

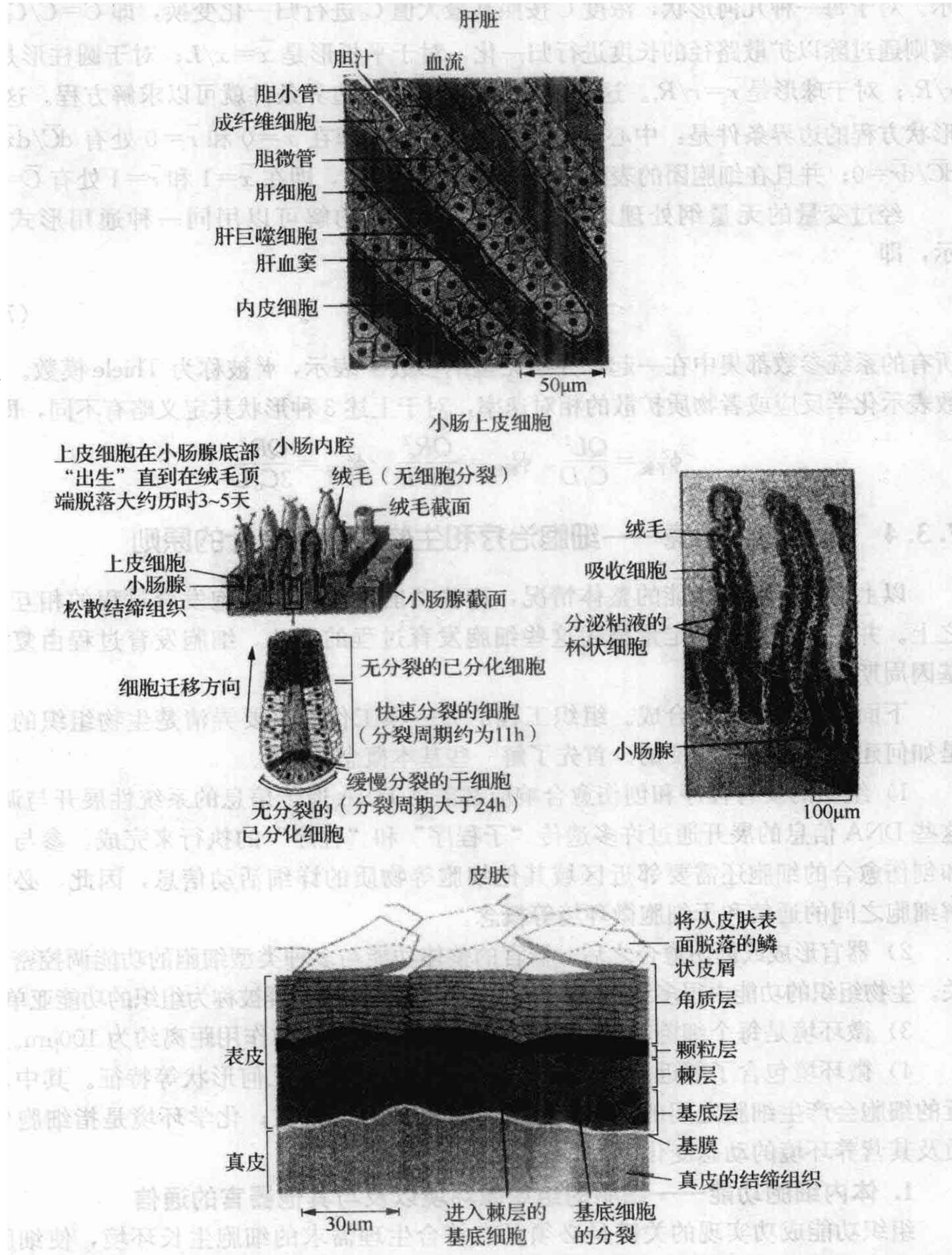


图 7.26 体内的组织微环境举例（摘自 Alberts 等，1994）

织微环境的例子，相关内容在 7.2.8 节已有介绍。

细胞外的信号通过细胞膜传递给细胞，并经历细胞质中的各种信号传导机制之后，才能够传递给细胞核。有些信号由各种可溶性生长因子传送，这些生长因子可以来自循环系统的血液或者来自各种相邻的细胞，它们与适当的受体结合之后就能够传递信息。营养物质、代谢废物和呼吸气体以高度动态的方式进出微环境。微环境的特性与细胞成分、ECM 和局部几何结构相关，而这些因素也同时为细胞提供重要的信息。微环境的尺度范围约为 100μm 数量级。在这种氛围中，细胞在各种调控机制作用下正常运作。如果这种氛围被破坏，那么，细胞就无法行使组织的功能。

例 7.5　请阐述正常和异常肝组织中肝星状细胞（Hepatic Stellate Cell，HSC）之间的相互作用以及细胞周围的微环境。

解：如图 7.27 所示，肝星状细胞是一种间充质细胞，位于肝细胞与肝血窦内皮细胞之间的 Disse 间隙中。正常的肝脏组织中，肝星状细胞紧贴基底膜样的基质，该基质由Ⅳ型胶原蛋白、层粘连蛋白和硫酸类肝素蛋白聚糖组成。如果肝脏受到损伤，肝星状细胞就会被激活，并开始生产基质蛋白，其中多数是Ⅰ型和Ⅲ型胶原蛋白。肝星状细胞的激活会使肝细胞和内皮细胞的表型都发生改变，也就是使肝细胞失去刷状缘，使内皮细胞失去窗孔。如果这种状态持续下去，肝脏就会纤维化。可见，组织微环境细胞功能调节的缺失会导致组织功能的丧失。■

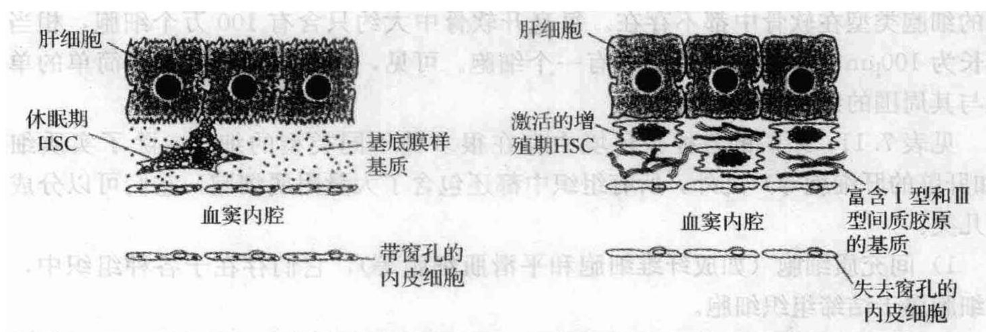


图 7.27　肝星状细胞的激活以及肝脏纤维化过程示意图（摘自 Iredale，1997）

组织受到微循环系统的灌流，其灌流在大于 100μm 的尺度上具有均一性。微循环系统与长距离的血液对流传输系统相连。对流传输系统连接人体的所有组织，给所有细胞提供营养、交换呼吸气体并排泄废物等。

各种组织都由功能亚单位组成，亚单位是组织的最小功能单位。从亚单位分离出来的细胞只具有细胞的功能，只有当多种细胞一起组成亚单位时才具有组织的功能。功能亚单位在不同物种之间具有相似性。每种组织所包含的亚单位数量取决于特定器官的整体需求。大象的肾脏肯定比小鼠的肾脏大得多，但是，两者的肾功能单位（即肾单位）却很相似。

在组织工程中，上述这些特性涉及两类问题，即宏观环境的设计和微环境的设

计。于是，组织工程的设计人员面临以下两大基本难题：

1）微环境的设计，包括其化学性质、几何特性、细胞结构和扩散特性等；

2）与其他各种组织的相互作用，包括营养的供应、呼吸气体的交换、废物的排泄以及生长因子等可溶性蛋白质分子的传输等。

模拟组织功能的装置必须能够提供图 7.2 所示的最外两层的功能，并且还要能够创造其他几层所表示的局部微环境。20 世纪 80 年代人们进行了大量生物反应器的设计工作，积累了很多经验，但是，那些反应器是培养单纯的连续细胞系的，用于蛋白质生产。显然，用于培养组织的生物反应器的设计会面临一系列新问题。下面将讲述这些问题，并介绍影响微环境的各种因素，包括微环境的细胞性、化学性质、动力学特性以及几何特性。

2. 微环境的细胞性

装在容器里的细胞密度约为 10 亿/mL 细胞的数量级，生物组织的细胞密度一般为该数值的 1/10～1/2，也就是 1 亿～5 亿/mL 细胞。微环境的尺度范围为 100μm，由此可以计算组织微环境中细胞数目的数量级。如果细胞密度为 5 亿个/mL，那么边长为 100μm 的立方体中约含有 500 个细胞。作为有趣的对比，在此容积下，简单多细胞微生物线虫（C. elegans）大约有 1000 个细胞。

不同组织微环境中的细胞性有所不同，细胞性最低的是软骨。软骨细胞的主要功能是维护细胞外基质。软骨无血管、无淋巴、也无神经。因此，很多其他组织含有的细胞类型在软骨中都不存在。每毫升软骨中大约只含有 100 万个细胞，相当于边长为 100μm 的立方体软骨中只有一个细胞。可见，软骨的微环境就是简单的单细胞与其周围的细胞外基质组成。

见表 7.11，组织的各种微环境中存在很多种不同类型的细胞。除了实质细胞（如肝脏的肝细胞等）之外，所有组织中都还包含了大量附属细胞，它们可以分成如下几类：

1）间充质细胞（如成纤维细胞和平滑肌细胞等），它们存在于各种组织中，这种细胞属于结缔组织细胞。

2）单核细胞在各种组织中都存在，它们具有独特的形态（见图 7.28）。单核细胞可以分化成为巨噬细胞，巨噬细胞一旦被激活，就会产生各种细胞因子和趋化因子，这些因子会影响邻近细胞的行为。

3）内皮细胞在所有组织中都与脉管系统相结合。这种细胞在运输细胞进出组织的过程中发挥着重要的作用，并且在组织的新陈代谢中也具有重要作用。

4）淋巴细胞和嗜中性粒细胞在组织中存在的时间很短，它们一般参与宿主的抵御反应或参与其他清除功能。

这些附属细胞一般占组织所有细胞的 30%，为了保持这个比例，实质细胞的数量会作相应的调整。表 7.12 列出了肝脏所包含的各种细胞的数据。

表 7.11　组织微环境中存在的各种细胞

基质细胞：由共同的前体细胞衍生而来
间充质细胞
成纤维细胞
成肌纤维细胞
成骨细胞/成软骨细胞
脂肪细胞
基质附属细胞：它们的产生过程与基质细胞不同，是组织的永久性居住者
内皮细胞
巨噬细胞
移行细胞：为了宿主的抵御反应，在炎性刺激前后迁移进入组织的细胞
B 淋巴细胞/血浆细胞
细胞毒性 T 细胞和天然杀伤细胞（即 NK 细胞）
粒细胞
实质细胞：占据组织大部分容积的细胞，它们表达组织的功能，并与其他种类的细胞相互作用，以便促进分化功能的表达

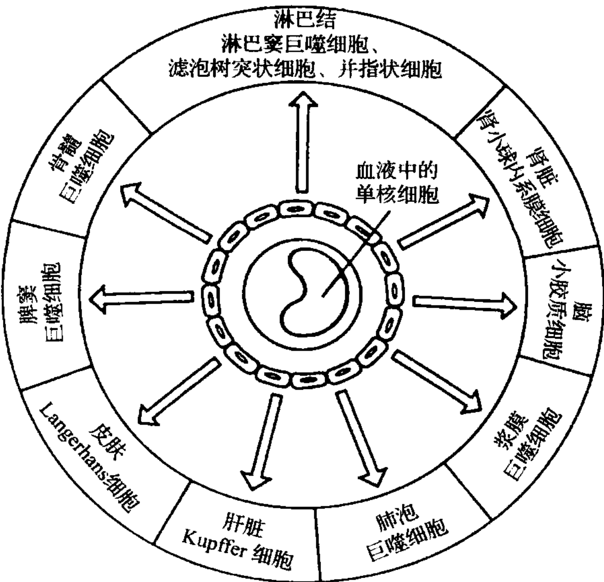


图 7.28　巨噬细胞在各种组织中的分布情况（摘自 Hoffbrand 和 Pettit，1988）

表 7.12　肝脏微环境中的各种细胞分布

细胞类型	大小/μm	占总细胞的百分比
基质细胞		
Kupffer 细胞	12～16	8
血管内皮细胞	11～12	9
胆管内皮细胞	10～12	5

（续）

细胞类型	大小/μm	占总细胞的百分比
贮脂细胞	14～18	3
成纤维细胞	11～14	7
陷窝细胞	11～15	1～2
实质细胞		
单核细胞（I型）	17～22	35
双核细胞（II型）	20～27	27
嗜酸性细胞（III型）	25～32	5

例 7.6 如果除掉组织中的附属细胞，那么会产生什么结果？

解： 有关骨髓培养中附属细胞（又称基质细胞）的作用已经有系统的研究。在体外培养实验中，根据已知的抗原可以将各种未成熟的细胞分离开来，从而可以改变实质细胞与附属细胞的组成比例，这就是附属细胞活性的滴定实验。如图 7.29 所示，实验结果表明，去除附属细胞之后，总细胞、祖细胞和前祖细胞等生产指标都急剧下降。补充经辐射预处理过的基质可以恢复总细胞数和祖细胞数，但不能恢复前祖细胞数。该结果与预期结果一致，也就是特异性附属细胞与实质细胞之间的相互作用对于未成熟细胞非常重要。■

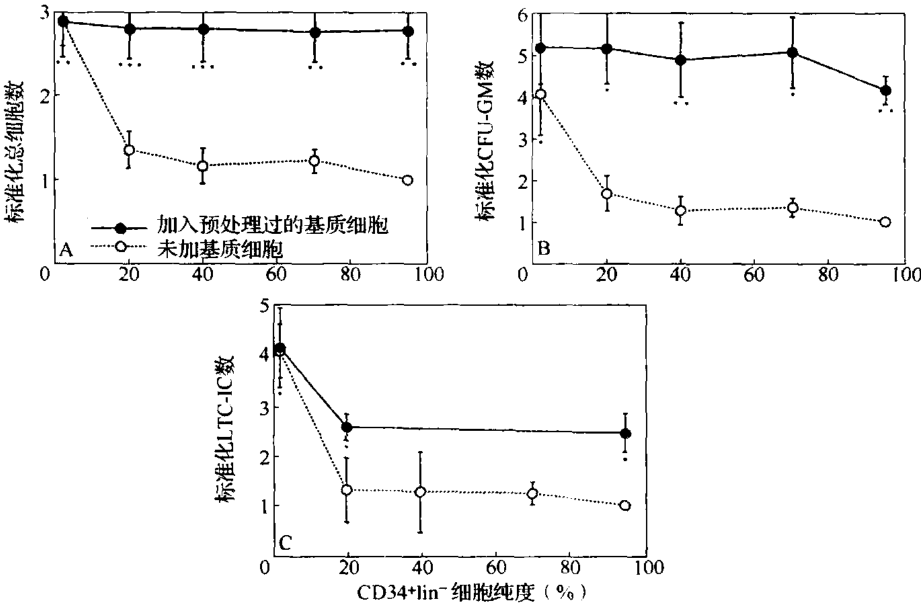


图 7.29 $CD34^{+}lin^{-}$早期造血细胞群的纯化对于细胞培养结果的影响（摘自 Koller 等，1995a）

注：纯化程度越高，由于缺乏附属细胞，有关细胞的生产指标会下降。其中，CFU-GM 是粒细胞/巨噬细胞集落生成单位，LTC-IC 是长期培养启动细胞。

3. 微环境的动力学特性

微环境具有高度动态性，需要用多个时间常数来描述。

(1) 氧化

一般，虽然哺乳动物细胞消耗氧气的速度比微生物慢，但是，如表 7.13 所示，就血液或培养基中的氧含量而言，它们摄取氧气的速率却很快。37℃温度下，水溶液的氧饱和浓度只有 210μmol/L。哺乳动物细胞的耗氧速率约为 0.05～0.5μmol/10^6细胞/h。如果组织的细胞密度为 5 亿/mL 细胞（即 5×10^8），那么，该耗氧速率所需的氧气输送速率为 25～250μmol/mL/h。体外组织培养时的氧气灌流速率和体内血液的氧输送速率都必须满足这个需求。

表 7.13　体外培养时各种人体组织和细胞的耗氧速率

人体组织或细胞	耗氧速率/(μmol/10^6细胞/h)
HeLa 细胞	0.1～0.0047
HLM（肝脏）	0.37
LIR（肝脏）	0.3
AM-57	0.045～0.13
皮肤成纤维细胞	0.064
Detroit-6 细胞（骨髓）	0.43
结膜	0.28
白血病 MCN	0.22
成淋巴细胞	0.053
肺	0.24
肠	0.40
二倍体胚胎 WI-38 细胞系	0.15
MAF-E	0.38
FS-4	0.05

由此可见，即使细胞密度很低，具有代谢活性的组织和细胞也会很快耗尽培养基中所含的氧气。例如，细胞密度为 10^6 个/mL 时，0.4～4h 内氧气就会被耗尽。因此，体外培养时，氧气必须连续供给。目前，肝细胞、角质形成细胞、软骨细胞和造血细胞等很多细胞都已经实现了体外培养，并用于临床细胞治疗。有人系统地研究了氧气对于肝细胞培养的影响，结果表明，肝细胞的耗氧速率（Oxygen Uptake Rate，OUR）约为 1.0μmol/10^6细胞/h，在各种哺乳动物细胞中属于较高的水平。相反，大鼠骨髓细胞的耗氧速率要低得多，只有 0.02μmol/10^6细胞/h。

除了氧气的供给以外，细胞附近的氧浓度必须保持在特定的范围内。氧化程度会影响很多生理过程，包括细胞粘附、细胞生长和细胞分化等。如果氧浓度过低，那么细胞的生长就会减慢；反之，如果氧浓度过高，那么，细胞的生长也会受到抑制，甚至会发生细胞中毒。例如，很多研究结果已经表明，5%的氧浓度可以使造血细胞集落的形成显著增加，另外，生物反应器中骨髓细胞的培养也存在最佳氧浓度。由此可见，细胞培养中氧气供给速率的设定非常重要。

(2) 新陈代谢与细胞信号传导

通常，虽然细胞对于营养物质的局部浓度有所反应，但是，细胞培养中营养物质的灌流要求并不高，因为培养基中营养物质的浓度可以比氧气浓度高很多，尤其是那些消耗速度较快的营养物质，可以配置高浓度。一般，葡萄糖的消耗速率约为 0.2μmol/10^6细胞/h，氨基酸的消耗速率为 0.1～0.5μmol/10^9细胞/h。不过，各种生长因子的输送速率和消耗速率却有较严格的要求。图 7.30 显示了几种物质的扩散响

应时间。

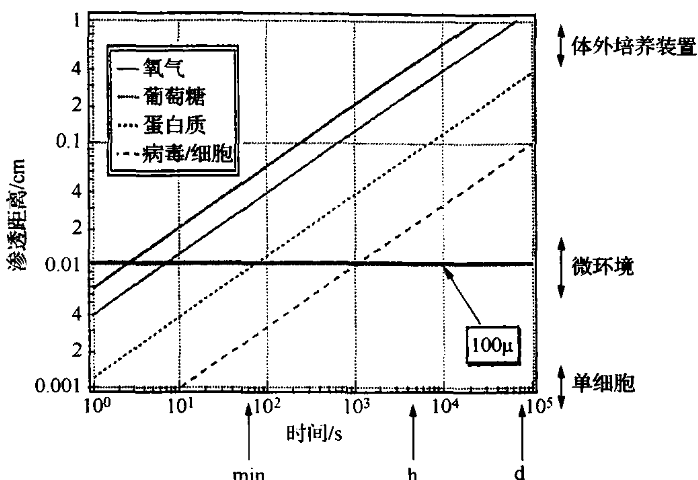


图 7.30 几种生物分子的扩散渗透距离随时间的变化

（3）灌流

人体循环系统将血流输送到各个器官，然后再分布到组织的微环境之中。人体总的灌流速率约为 5L/70kg/min，相当于 0.07mL/CC/min。如果每 ml 组织中含有细胞 5 亿个，那么，该灌流速率就相当于 0.14μmol/10^6 细胞/min。这些数据只代表整个人体的平均值。不同器官的灌流速率有所不同，它与器官的代谢活性有关。

（4）细胞迁移

正如前面章节所述，细胞会运动，并且各种细胞的运动速度各不相同。嗜中性粒细胞在一分钟内就可以移动几个细胞直径的距离，而角质形成细胞、纤维细胞和内皮细胞等粘附细胞却一小时只能移动一个细胞直径的距离。就细胞复制和细胞分化的速度而言，细胞迁移属于一种快速过程。嗜中性粒细胞的迁移是对微生物入侵做出的快速响应，而粘附细胞的迁移则是对组织的动态需求做出的响应。

（5）基质更新

细胞外基质时刻都在更新。在心脏中，每天约有 3%的细胞外基质得到更新。

4. 微环境的几何形状和大小

（1）几何形状

如图 7.26 所示，各种微环境的几何形状各不相同，且它们的维数也不同。许多微环境呈现二维有效维的曲面形态，而骨髓中的细胞排列则表现出分形几何特征，其有效维数为 2.7，大脑的有效维数则是三维。

（2）大小

什么因素决定了微环境的大小？这个问题至今仍然没有明确的答案。其最小尺寸应该就是单细胞的大小，约为 10μm，细胞团的大小应该是这个尺寸的倍数。决定微环境最大尺寸的因素尚不明确；不过，生长因子的有效信号传播距离以及氧气渗

透距离的实验数据表明，细胞之间的通信和细胞代谢速率的动态特性在决定微环境尺度上具有重要的作用，而这些距离都取决于物质的扩散过程，它们的估计值约为100～200μm。另外，随着细胞数量的增加，各种细胞功能调控的稳定性问题也会凸现出来，可能成为实际应用中的一个限制条件。

7.3.5 生物材料

组织工程中的生物材料研究仍然存在一些难题，下面从 3 个不同尺度上进行讨论。最小的尺度在生物化学层次上，这里的问题包括细胞外基质 ECM 的化学特异性及其与各种细胞受体之间的相互作用（参见 7.2.8 节）。可以将完整的 ECM 成分涂在材料表面，以确保细胞与其密切接触的环境之间具有适宜的相互作用；但比较精细的处理方法是合成 ECM 中存在的特异性键合蛋白序列，并通过某种方式呈现给细胞，例如，通过将这种蛋白质制作成微型图案，就可以获得特定的细胞排布。解决这些问题需要材料制造技术、生物化学和遗传学等多个领域的研究人员的共同合作。

第二个尺度在 100μm 数量级，也就是器官微环境的尺度。很多器官具有高度特异性的局部几何特征，体外培养系统也应该模拟这些特征，建立具有特定力学特性的微结构。显然，这给生物材料的制造提出了难题。而且，基质支撑材料应该具有生物可降解性，植入体内之后，可以被降解，且降解产物必须无毒性。聚乳酸类和聚乙醇酸类聚合物是发展前景很好的生物降解性材料。另一方面，如果移植之后，植入体不发生什么重塑，那么，基质支撑材料的宏观几何尺寸就很重要。

最大的尺度就是生物反应器本身。组织工程中使用的生物反应器其实都很小，其尺寸只有 10cm 数量级。这个尺度下的材料问题主要是生物相容性。虽然现在已经有用于组织培养的塑料产品，但是仍然存在一些问题。例如，商品化的塑料培养器皿一般设计成能够促进细胞的粘附、键合以及连续细胞系的传代，这些性质对于连续细胞系的培养比较理想，但对于各种原代细胞培养却不一定有利。

7.4 体外培养中的工程学问题

7.4.1 基本原则

如前所述，生物组织的动力学特性包含了细胞复制、细胞分化以及细胞衰老等一系列细胞发育过程之间的复杂相互作用。在体内环境中这些过程处于平衡状态，体内的动态平衡是各种生物动力学过程和物理化学的制约共同作用的结果，因此，在体外重建生物组织时必须建立生物学和物理化学的动态平衡，这样才能够实现正常的组织功能。

7.4.2 关键设计问题

根据上述原则，下面我们通过组织重建成功的例子来了解组织工程中的一些问

题，这里只作简单介绍，并不是全面列举组织工程中的重要问题，也不是所有已发表文献的全面综述。我们的目的是阐明组织工程中的工程学问题，并明确体外人体组织重建中工程学所起的关键作用。

组织工程中的几个重要设计难题是：

1）氧化，即提供符合生理浓度的氧气流量；

2）提供或者去除细胞因子和趋化因子；

3）各种生理物质的灌流速率和物质分布的均匀性；

4）生物材料的功能、结构和毒性等性质以及制造技术。

下面将简要介绍这四大问题。除了这些问题之外，临床细胞治疗中相关的其他问题还有：一次性器械的设计、培养基组成成分的优化、初始细胞排布、满足 FDA 的要求以及临床操作方法等。这些问题不在此详细介绍，不过，以下论述中会涉及其中的某些问题。

7.4.3 物质传输的时间问题

生物组织和细胞功能的实现中物质传输的重要性经常被忽视。如图 7.30 所示，如果根据扩散系数作图显示与组织功能相关的各种物质的平均扩散距离随时间变化的曲线，那么，分子扩散的局限性就显而易见。在生理学时间尺度上，物质扩散可以渗透的距离短得惊人，这种渗透特性决定了人体各种器官的结构。体外培养装置面临着同样的物质传输约束条件，具有生理细胞密度的细胞床中很难获得较高的物质传输速率。

微环境的生化特性对于正常组织功能的形成非常重要。有关连续细胞系生长所需的生化条件已有很多现成的数据可以查询，连续细胞系的体外培养始终围绕着营养物质供给和废物排泄的问题。原代细胞培养时，营养物质还有其他作用，会直接影响细胞的生理性能。例如，最近有研究表明，脯氨酸浓度和氧浓度在肝细胞培养中具有重要作用。

在多数情况下，氧气的输送是重点需要考虑的问题。氧气过量会抑制细胞的生长或者引起氧中毒，而过少又会影响细胞的新陈代谢。肝脏、肾脏和脑等组织的需氧量较高，其他组织的需氧量则较低。要将氧浓度控制在理想的水平上，达到所需的均一性，并且还要在高细胞密度下维持所需的氧流量，这确实很困难。而且，在重建包含多种不同细胞的生物组织时，各种细胞的氧气和营养物质需求量都不同，这使得物质输送变得更加复杂。由此可见，体外培养中微环境生化特性的确定、设计和控制都十分困难，尤其在物质扩散效应和特定的微结构条件限制下，其难度更大。

例 7.7 如何测定培养物质的耗氧速率？

解： 多数耗氧速率的数据都是指悬浮细胞培养的耗氧量，它们用标准呼吸计测定。但是，精确的原代细胞培养的耗氧量必须在贴壁生长状态下测定，如图 7.31 所示是一种适用于这种测定的新型呼吸计。用这种呼吸计测得的肝细胞培养的耗氧速率（即 OUR）为 1.0μmol/10^6 细胞/h；而类似仪器测得的骨髓细胞培养的 OUR 数值

为 0.03～0.04μmol/10^6细胞/h，这个数值接近体内细胞的耗氧速率。 ■

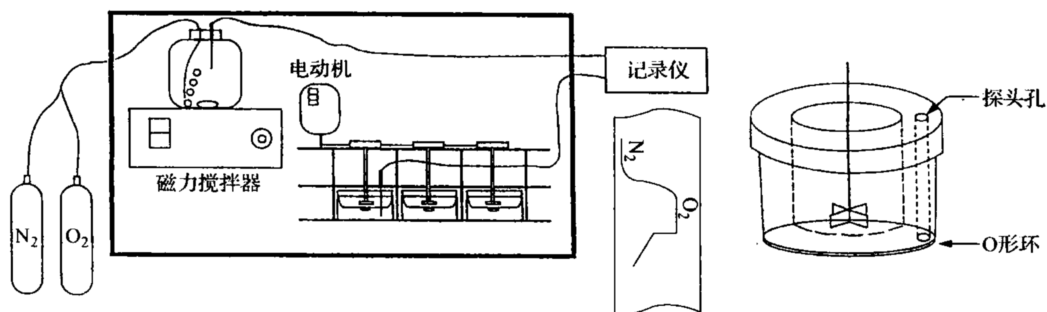


图 7.31　耗氧速率测量装置示意图（摘自 Peng 和 Palsson，1996a）

例 7.8　如何输送氧气?

解：如图 7.32 所示，氧气可以通过氧化膜来输送。当灌流液的流速较慢时，其中所含的氧气在入口附近很快就会被耗尽，因此，细胞孵育床中的供氧主要来自氧化膜中氧气的侧向扩散。这样，只要调节气相中的氧含量，就可以控制输氧量。这种输氧方式可以独立调控输氧量，与其他参量无关，并且输氧的均匀性很好。 ■

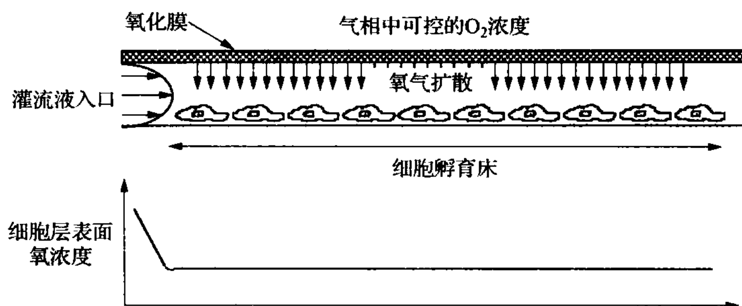


图 7.32　通过氧化膜输送氧气的方法

注：氧化膜覆盖在细胞孵育床灌流液的上方，如果灌流速度较慢，那么，孵育床中的氧气主要来自氧化膜上的氧扩散。这样，只有在灌流液入口处的小范围内氧浓度受到灌流液的影响，孵育床其他区域的氧浓度很稳定。

7.4.4　液流及其均匀性

微环境的尺度为 100μm。体外培养装置包含有数千万个微环境，它在输送和排除气体、营养物和生长因子时必须保持均一性，不过，这种均一性很难实现。困难在于液体流过固体表面时会产生无滑移现象，也就是，在生物反应器内壁附近总是存在流速较慢的区域，这些区域的轴向 Graetz 数小于 1，从而导致这些区域的物质扩散量超过其他区域。

图 7.33 说明了上述问题的起因。图中所示是液体顺着狭缝流下的情景，图 7.34

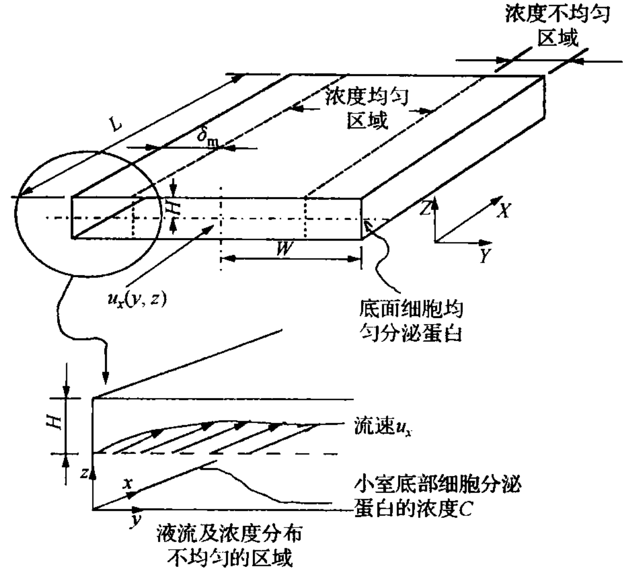
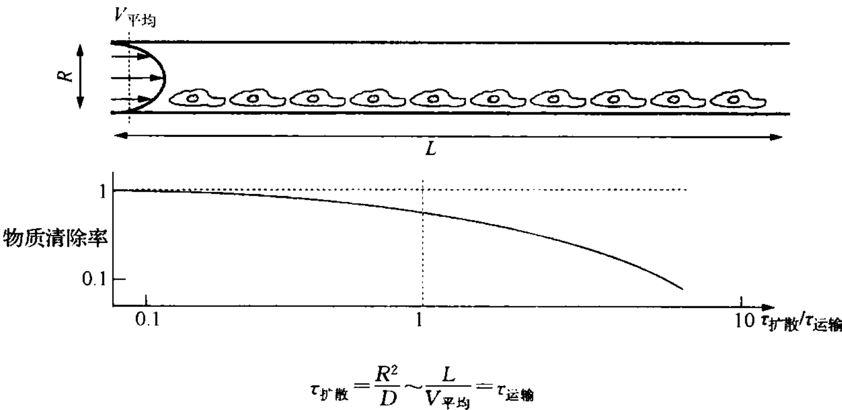


图 7.33　矩形狭缝小室坐标系示意图（摘自 Peng 和 Palsson，1996b）

注：其中位于底部的细胞群会分泌蛋白因子。侧壁附近液体的流速较慢，使这些区域的物质浓度与狭缝中间区域的浓度不同，从而在沿壁区域形成了与其他区域完全不同的细胞生长和组织功能发育的微环境。



$$\tau_{扩散}=\frac{R^2}{D}\sim\frac{L}{V_{平均}}=\tau_{运输}$$

图 7.34　液流中生长因子浓度分布示意图

注：当液体流过细胞孵育床时其中的生长因子会被细胞吸收。下图：液流流过细胞床之后其生长因子清除率随两个时间常数之比而变化的曲线，其中，$\tau_{扩散}$为生长因子侧扩散的时间常数；$\tau_{运输}$为生长因子随液流运输的时间常数。这两个时间常数之比就是 Graetz 数 Gz。如果 $Gz>1$，那么，生长因子蛋白分子就没有足够的时间扩散并与细胞结合，大部分生长因子就会从出口处流出；反之，如果 $Gz<1$，也就是扩散时间远小于运输时间，那么，生长因子就有充裕的时间与细胞结合，被细胞吸收，于是，流出孵育床的生长因子就很少。

显示了狭缝的剖面图。宽度与高度之比很大的窄缝中的流体可以近似看作 Hele-Shaw 流，除了两侧边缘附近之外，其他区域内的液流基本相同。两侧流速较慢区域的宽度与狭缝的高度相当。因此，如果狭缝的宽度与高度之比为 10：1，那么，在整个狭缝宽度的 90%区域内液流都可以看作均匀一致；而在其余 10%靠近侧壁的区域内较慢的液流会产生不同的细胞生长微环境，从而造成不同的细胞生长速度，并促使其他区域的细胞向侧壁迁移。实验结果已经证实了这种内壁附近非均匀细胞生长现象的存在。使用放射状灌流的方法可以解决这个问题。

以上例子只是细胞培养生物反应器的一种结构，其他结构的装置也会面临同样的均一性问题，都必须仔细分析物质的动力学传输过程，以便指导具体装置的设计。

7.5　组织工程化产品的临床应用问题

上一节介绍了在体外培养临床应用所需的细胞数量时组织工程技术人员所面临的设计难题，这些问题只是细胞治疗实施过程中所遇到的诸多问题中的一部分，本节将介绍其他必须解决的一些难题。

1. 供体之间的差异

人体的遗传变异性很大，这使得不同供体的体外细胞培养结果差别极大，某些关键指标的标准差可以超过平均值的 50%；因此，即使细胞生长的处理过程、使用的材料、培养基的配方等都一样，细胞培养结果仍然会相差很大，需要经过多次实验（10～12 次）才能得到满意的结果。

这种变异性是生物体内在固有因素决定的。补充足够的附属细胞可以调控细胞培养的结果。虽然供体之间存在明显的变异性，但是，相同组织源的行为具有内在的一致性。图 7.35 所示是几种不同供体样本的生长因子吸收速率之间的相对关系，可见，这些生长因子吸收速率之间存在很高的相关性；但是，这种相关性的定量描述在不同供体之间的差别很大，很难建立适用于较大供者群体的关系方程。

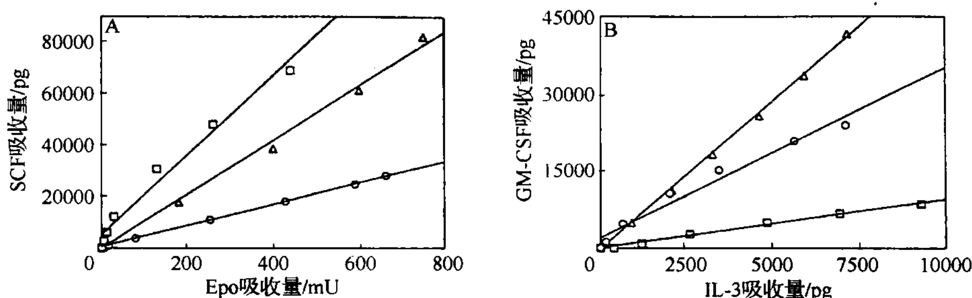


图 7.35　各种生长因子吸收量之间的关系（摘自 Koller 等，1995b）

注：左图：人体骨髓培养中干细胞因子 SCF 与促红细胞生成素 Epo 吸收量之间的相关性；右图：粒细胞—巨噬细胞集落刺激因子 GM-CSF 与白细胞介素 IL-3 之间的相关性。对于同一个组织样本，这些相关性很强；但是，不同供体样本的拟合直线的斜率差别很大。

2. 各种参数之间的密切联系

原代细胞培养对很多参数都很敏感，因此，可以通过设置这些参数来调控细胞的培养结果。不过，即使不是所有参数，很多参数之间都存在密切联系，如果细胞培养中的某个参数发生了变化，它就会影响其他参数的最佳数值。利用统计学实验设计方法可以快速搜索实验中所涉及的多个参数。图 7.36 显示了一种二维搜索的方法，可见，人体细胞培养中祖细胞的最佳产量是细胞接种密度和培养基更新速率的函数。

图 7.36 中的两张图使用了不同的计算指标，左图所示是细胞生长单位面积上的祖细胞产量，该指标的最优数值可用于指导所需数量细胞培养装置的最小化设计。右图所示是祖细胞的扩增率，也就是输出细胞数与输入细胞数之比。当初始细胞数量有限，且希望获得最大的祖细胞增殖时，可以根据这个指标进行设计。注意，这两个指标的适用情况有所不同，因此，在一开始就必须明确所需分析的计算指标，但有时这一点却很难确定。

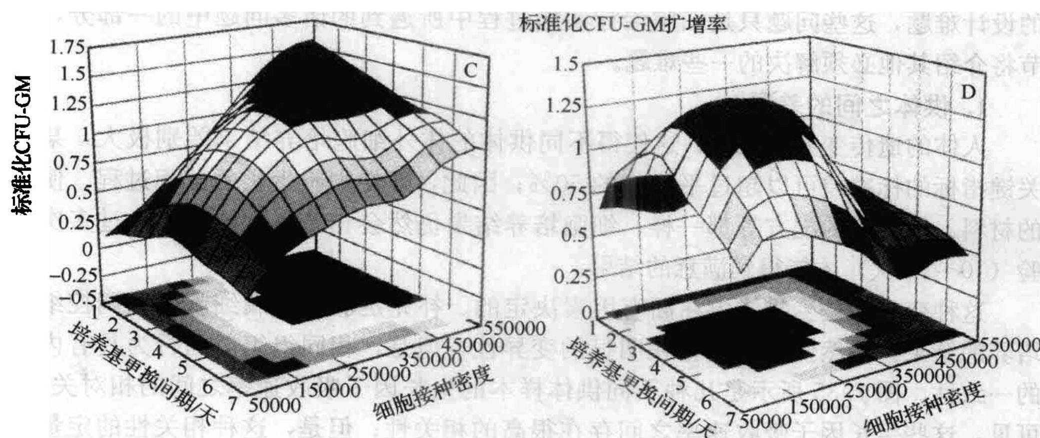


图 7.36 不同细胞接种密度和不同培养基更换间期下骨髓细胞的培养结果（摘自 Koller 等，1995c）

注：注意各个变量之间存在相关性。其中，左图的 z 轴表示单位面积的 CFU-GM 祖细胞总数，右图的 z 轴则表示细胞的扩增率，也就是输出细胞数与输入细胞数之比。注意，这两个指标最优化所对应的细胞培养状态完全不同。

3. 免疫排斥

异基因移植都会遇到免疫排斥问题，不过，排斥情况有所不同。例如，用于治疗皮肤溃疡的真皮纤维细胞移植没有致免疫性，其原因尚不明确。这样，利用同一供体的原料就可以培养许多移植体，用于多位病人的移植。相反，胰岛 β 细胞却具有排斥性（见图 7.6）。而且，异基因骨髓移植时，移植体还会排斥免疫缺陷的宿主；因此，移植物抗宿主病是引起异基因骨髓移植致死的主要原因。

人们对免疫反应的细胞机制和分子机制的了解正在不断深入。排斥反应是 $CD4^+$ 细胞（CD4 是 T 细胞的一种表面抗原）的两个细胞亚群 Th1 与 Th2 之间相互作用的

动态过程，Th1 和 Th2 的细胞因子分泌特性和功能都有所不同。参与免疫调节的各种成分和机制也正在清晰起来。免疫系统的定量分析将有助于制定预防和治疗排斥反应的合理方案。

4. 组织的获取

体外培养的材料来源非常重要。用于皮肤置换的真皮纤维细胞一般取自人体包皮，通过包皮环切手术获得。这种细胞的增殖能力很强，可用于生产大量移植体。由于来源相同，细胞培养过程中的生物变异性（参见 7.5.1 节）大大减小，这样，最终产品的成本中组织获取的费用就可以降低。相反，成人自体移植材料的获取费用可能会很高，并且，还会表现出很大的变异性。不过，用于细胞治疗的最理想的材料是组织特异性的干细胞，其自我更新速率和分化的定型都可以调控。而且，如果干细胞移植体可以变成无致免疫性，那么，多数与组织获取相关的问题也就迎刃而解了，组织生产过程中的变异性也可以减小，产品的质量保证和质量控制也就变得很简单。

5. 低温贮藏

低温贮藏是生物物理学、化学和工程学的基本物质存储方法。如今，这种方法在临床上已用于骨髓、血细胞、角膜、生殖细胞和血管组织等的贮藏。但是，最近却有人发现，低温贮藏一般不能用于体外培养的人体细胞，因此，亟待开发新的方法用于这类产品的贮藏。现在，涉及体外培养的细胞治疗中，低温贮藏只能用于细胞培养之前原始组织的冷冻保存。

7.6　未来的发展方向——功能性组织工程

7.6.1　细胞水平的发展

某些用于移植的细胞群体中可能含有不利的细胞亚群。例如，自体采集的造血细胞中可能含有患者的肿瘤细胞。理论上，移植之前应该清除这些污染细胞。同样，活组织检查时也会受到附属细胞污染的影响，附属细胞的生长速度可能比待检测的实质细胞还要快。例如，许多活检中成纤维细胞就是一种很难清除的污染细胞，在培养过程中，这种细胞的生长速度往往特别快。正是由于这个原因，很多肿瘤组织的原代细胞系很难培养。

7.6.2　功能性组织工程

组织工程学与细胞生物学的区别就在于它注重组织功能的实现，也就是将大量细胞组成高度有序的结构，根据解剖复杂性和结构的集成性，形成组织或者器官。肢体和器官系统等复杂的人体解剖结构的重建是一个系统工程问题。虽然有人会辩解说，组织的所有功能都来自基本的细胞机制，但是，支架中的一个个细胞或者没有组织性的细胞团不可能产生具有系统水平的各种组织和器官的功能。这就像一堆砖头没有房屋的功能、一箱汽车零件不具有汽车的功能一样。同理，工程化的生物

组织必须在系统水平上进行考察，工程工作的成功与失败最终要用产品在器官水平上功能的定量评估来决定。显然，应用分子生物学技术验证一种或多种主要亚细胞组分的存在并不足以证实最终的器官功能。因此，必须开发必要的工具和方法来定量检测组织工程产品的器官水平的功能。当然，不同的组织和器官具有不同的功能。由此可见，在设计组织或器官产品时，必须制定具有明确定义的产品的技术规格、功能评价的定量指标（即品质因数）以及检测这些指标的具体方法。

可以推测，组织或器官产品在使用之前都要经过一段时间的培养，以便得到生长发育。很多组织的功能会在发育过程中产生变化，因此，最好选定一个或几个能够在组织培养过程中进行非破坏性检测的参量。举一个值得借鉴的例子：从发育早期到成年，哺乳动物骨骼肌的收缩性能都一直在变化。肌肉的性能主要取决于构成肌纤维的肌球蛋白重链，通过肌肉组织的等长收缩实验可以无破坏地测定组织的性能。肌腱也是如此。在发育过程中，肌肉组织的切线模量、抗拉强度和断裂韧度都会逐渐增加，而应力应变曲线的斜率却会逐渐减小，这可能是由于出生前后发育时期胶原蛋白结构的有序性增加的结果。对于体外培养的肌腱样本，不可能无破坏地测量其抗拉强度和断裂韧度，但是基本可以无破坏地测量其切线模量和应力应变曲线的斜率。对于肌肉骨骼系统的组织，进行无破坏定量测定时，经常需要外加电刺激信号，这就像组织培养过程中用电刺激诱导和促进肌肉的发育一样。目前已经使用电刺激诱发骨骼肌细胞和心肌细胞收缩的方法来促进其发育。早在20世纪80年代人们就开始对肌肉骨骼系统的各种组织（包括肌肉、骨骼、肌腱、软骨和韧带）和血管组织施加机械应变，从而促进体外培养组织的发育。

未来亟待解决的难题是开发一种生物反应器系统，使其既可以用于施加电刺激信号，同时又可以无破坏地测量和记录组织的各种功能特性。如果生物反应器系统中能够实时监测正在发育的组织的功能特性，那么就可以获得有关组织样本当前发育状态的信息，并根据这些信息调整刺激参数。这样就可以在组织发育过程中实现刺激信号的反馈控制，从而显著提高功能性组织培养的工程技术水平。这是目前肌肉骨骼系统和心血管系统功能性组织培养领域的一个重要研究方向。

7.6.3 生物人工肝

生物人工肝（Bio Artificial Liver，BAL）的发展源于治疗人类肝病的社会需求。人体的肝脏系统没有“后备”部件，不像肺脏、肾脏等都有2个；因此，具有维持人体内稳态多种重要功能的“单个”肝脏，一旦发生暴发性肝功能衰竭（Fulminant Hepatic Failure，FHF）等疾病，就需要用到替代系统。

有关肝病治疗的报道最早出现于20世纪50年代，当时提倡用低蛋白饮食缓减精神障碍和肝性脑病。60年代出现了肝脏辅助器械的新概念，于是，各种替代特异性肝功能的治疗方法从此发展起来了。目前肝脏人工辅助系统（Artificial assist Systems，Artif-S）有的仍在研究中，有的已具备治疗FHF等肝脏疾病的功能，并已进入FDA试验初期。其中包括用于氨清除的炭过滤器、用于排毒的机械透析装置以及

用于清除血液循环系统中病变物质的血浆置换器等技术。研究结果表明，许多人工辅助系统都能够实现某些特定的功能，但是，由于它们能够完成的功能很少，因此，并不能彻底解决问题，不能完全替代肝脏的功能。根据市场的需求，人工辅助系统一直在不断改进，如肾透析机使用了先进的设计参数等，但是，这些机械装置仍然无法复制健康人体肝脏能够完成的众多功能。

器官移植是能够成功治疗 FHF 疾病的一种方法，它用能够完成所有代谢功能的健康肝脏代替丧失功能的肝脏。成功的移植手术可以使病人摆脱使用机械人工辅助装置的困扰。其缺点是病人必须服用免疫抑制剂来减轻排斥反应，不过，这并不是大问题。但是，即使移植手术可以成功，肝源供体却很有限，并且必须满足组织配型条件，这致使肝源需求量大约高于供应量的 300%。因此，急需开发一种能够弥补这个供体缺口的方法。

于是，有关研究工作从人体肝脏组织原材料又回到了工程化的替代品。这个回复是目前科学技术发展的结果，它将迎来未来肝病治疗的成功应用。如今，FHF 的治疗仍然受到人工辅助系统功能的不足以及器官移植的供求失衡这两方面的限制，这表明某种活体组织与人工材料的结合可能是未来医学治疗的一个发展方向。

7.7 总结

目前正在开发的许多细胞治疗新方法给工程化组织培养提出了很多难题。要解决这些问题，首先必须了解组织的生理功能，定量描述细胞发育过程及其调控机制；然后，设计各种物理化学参量的变化率，使它们能够符合细胞发育的需求，并形成组织功能。组织工程学的发展目前尚处于“胚胎期”，各种参量的数量级和维数的定量分析对于组织功能的设计和重建具有重要意义。毫无疑问，在未来几十年中工程化生物组织的性能将显著提高。

7.8 术语

同种异体移植（Allogeneic）：同一物种但不同血统的供者的细胞或组织移植给受者。

凋亡（Apoptosis）：细胞衰老并死亡的过程，这种过程由细胞本身引起。

自体同源移植（Autologous transplant）：取自生物机体的细胞或组织再植回该生物体的移植。

BMT（Bone Marrow Transplantation）：骨髓移植。

细胞治疗（Cellular therapies）：利用人体的早期细胞治疗疾病的方法。

软骨细胞（Chondrocytes）：存在于软骨中的细胞。

集落形成测定法（Colony-forming assay）：半固体培养基中在生长因子的刺激作用下，祖细胞产生分裂，生长出来的后代细胞聚集在一起，两周后用显微镜就可以识别形成的各个细胞集落。

细胞分裂（Cytokinesis）：细胞合成DNA之后产生的完整分裂，该过程使一个细胞分为两个。

细胞分化（Differentiation）：细胞或细胞群体向更成熟状态发展的不可逆过程。在分化过程中，各个细胞表达的基因及其表达水平各不相同。

移植期（Engraftment）：骨髓移植后维持血液循环中成熟血细胞安全数量的时期。

细胞外基质（Extra Cellular Matrix，ECM）：存在于细胞之间的不溶性蛋白质和碳水化合物的复合物，其作用是：在细胞群体之间建立物理连接，并通过特异性方式向细胞传递信号。

流式细胞仪（Flow cytometry）：通过荧光标记单克隆抗体来识别某些细胞类型的分析技术。较精密的流式细胞仪能够将被分析的细胞分成各种不同的细胞群体。

功能亚单位（Functional subunit）：器官的最小功能单位，如肺脏的肺泡、肾脏的肾单位等。

移植物抗宿主病（Graft-versus-host disease）：移植细胞对抗新宿主组织的免疫反应。异基因骨髓移植时这种反应经常会产生严重的后果，甚至导致死亡（即急性移植物抗宿主病）或终身残疾（即慢性移植物抗宿主病）。

造血（Hematopoiesis）：通过多细胞系的增殖和分化调节成熟血细胞的生产过程。

细胞增殖（Hyperplasia）：完整的细胞分裂生长过程，既包含DNA的合成，也包含细胞的有丝分裂。

细胞肥大（Hypertrophy）：只有DNA合成而没有细胞分裂的细胞生长过程，其结果是形成多倍体细胞，并使细胞的体积变大。在细胞分化的晚期，很多组织细胞系的细胞在再生刺激的作用下会变成肥大细胞。

细胞谱系（Lineage）：产生某种特定种类的成熟细胞过程中所有分化阶段的各种细胞之间的宗谱关系。

长期培养启动细胞（Long-Term Culture-Initiating Cell，LTC-IC）：在体外7～12周培养后得到的细胞。LTC-IC细胞被认为是非常早期的细胞，其中包含有干细胞，但不全是干细胞，它是多种不同细胞的混合群体。

间充质细胞（Mesenchymal cell）：成纤维细胞、成骨细胞、软骨细胞、脂肪细胞等结缔组织细胞。

微环境（Microenvironment）：体内细胞周围的环境。

有丝分裂（Mitosis）：导致细胞分裂的过程。

单核细胞（Mononuclear cell）：将骨髓离心之后获得的其中一种细胞，它与无细胞核的血红细胞和具有多形核的粒细胞不同。

骨髓抑制（Myoablation）：红细胞、白细胞和血小板等来自骨髓的各种细胞都死亡的状态，如接受高剂量化疗的病人就会发生这种情况。

实质细胞（Parenchymal cell）：器官组织独特的主细胞，如肝脏的肝细胞、肌肉组织的肌细胞等。

祖细胞（Progenitor cell）：由干细胞演变而来的单能细胞，它将继续分化为成熟细胞。

自我更新（**Self-renewal**）：产生与母细胞具有相同特性的子细胞的过程。常指干细胞分裂形成新的干细胞的过程。

干细胞（**Stem cells**）：能够自我复制的多能细胞，它具有无限增殖的潜力。恶性肿瘤细胞就是一种异常干细胞。

基质细胞（**Stromal cells**）：与上皮细胞在一起的间充质细胞。它们具有年龄特异性和组织特异性，它们的作用是调节上皮细胞的扩增和分化。例如，骨髓培养时所需的所谓骨髓基质层就是各种附属细胞和支持细胞的混合物。

同基因移植（**Syngeneic**）：遗传物质相同的供者与受者之间的细胞或组织的移植。

异种移植（**Xenogeneic**）：不同物种供者与受者之间的细胞或组织移植。

习　题

1. 请根据以下数据判断人体造血干细胞是否确实能够在体内进行自我更新：

- 每天大约生产 4000 亿成熟的造血细胞；
- Hayflick 界限的最佳估计值是 44～50；
- 大约 1/5000 的祖细胞不发生凋亡；
- 每个祖细胞大约发生 6～10 次倍增，共生产 50～1000 个成熟细胞；
- 整个分化途径可以包含 17～20 次倍增过程；
- 体外培养时单个高度纯化的造血干细胞（或前体祖细胞）最多可以产生 1～3 千万个细胞。

请给出推算过程，并分析各个参量对于造血过程的调控作用。其中调控端粒酶活性的参量在决定人一生中生产的成熟后代细胞的总量中具有什么重要作用？

2. 假设某种溶液浓度为 1pmol/L，那么，体积等于细胞大小（半径为 5μm）的溶液中含有多少个分子？

3. 请利用式 7.3 表示的连续方程，证明有复制无凋亡（即 $\alpha=0$）的细胞分化过程达到稳态时的细胞数目为

$$\frac{X_{\text{out}}}{X_{\text{in}}}=e^{\mu/\delta}$$

可见，它是比值 μ/δ 的函数。式 7.3 中参数 a 表示细胞分化的状态，$a=0$ 表示尚未分化的状态（细胞数为 X_{in}），$a=1$ 则表示分化结束的状态（细胞数为 X_{out}）。如果细胞生长速率 μ 和分化速率 δ 的数值大小相当，那么，分化结束时的细胞数目 X_{out} 会怎样？如果 δ 比 μ 小 10 倍，那么，X_{out} 又会变成怎样？这两种 δ 与 μ 的比值中哪一个在生理学上更合理？注意：如果这两个速率相当，那么每个原代细胞就只能生产 2 个后代成熟细胞。另一方面，如果细胞分化速率比增殖速率小 10 倍（这种情况与很多生理状况相似，也就是细胞倍增时间约为 20h，而细胞分化时间约为 200h＝8 天），那么，每个原代细胞就可以产生大约 10^6 个成熟细胞。由此可见，生物组织的整体动态过程取决于细胞发育的各个速率之间的相对关系。

4. 请按照如下步骤建立描述细胞分化的动力学连续模型：

(a) 推导描述细胞群体的一阶偏微分平衡方程；

(b) 利用分化速率 δ 将时间变量无量纲化；

(c) 利用分化速率 δ 将细胞凋亡速率和细胞增殖速率分别用无量纲参数 A 和 B 表示；

(d) 当 $A=0$ 时求方程的稳态解；

(e) 当 A 的取值为 $a_1 \sim a_2$ 之间的某个非零值时（即部分分化时），求方程的解；

(f) 当 $A=0$ 时求方程的暂态解。

5. 假设某个平面上有 2 个细胞，其中一个细胞向另一个细胞分泌趋化因子。请推导如下趋化因子浓度的稳态分布方程：

$$C(r)=\frac{R^2F}{D}\cdot\frac{1}{r}$$

式中 R——细胞半径；

F——分泌速率，即单位时间单位面积分泌的分子数量；

D——趋化因子的扩散系数；

r——离开细胞表面的距离。

如果不存在自由运动（即 $\mu=0$），请利用细胞的流量方程计算响应趋化因子的细胞迁移到信号细胞处所需的时间。有关参数值如下：

$\chi(C)=20\text{cm}^2/\text{s}\cdot\text{M}$；$D=10^{-6}\text{cm}^2/\text{s}$；$R=5\mu\text{m}$；

趋化因子分泌速率＝5000 个分子/细胞/s。

提示：将流量 J 的方程简化为 $v=\chi dC/dr$，其中 v 是细胞迁移速度。

6. 假设浓度为 C 的某种溶液从容器表面的一个直径为 d 的圆形小孔流出，扩散进入另一种溶液中，那么，其分子扩散流量 F 可以计算如下：

$$F=4DdC$$

总的流量就等于单孔流量乘以小孔数目。如果扩散系数 $D=10^{-6}\text{cm}^2/\text{s}$，浓度 $C=1\text{mmol/L}$，小孔直径 $d=1.5\text{nm}$；请计算单个小孔的流量。如果小孔代表两个相邻细胞之间的连接通道，那么，要达到有意义的信号通信，需要多少小孔？假设这是两个上皮细胞，且其形状近似方形，就像两个并在一起的方盒子，盒子之间有 n 个小孔相连，那么，在上述单孔流量下，请推导代谢物从一个细胞流向另一个细胞的时间常数计算公式。

7. 如果软骨组织中细胞的密度为 10^6 个细胞/mL，请计算细胞之间的平均距离，并分析这种微环境的特性。

8. 利用一维模型分析氧气向贴壁生长的细胞层的输送量，请证明例题 7.7 中的最大单位面积输氧量为

$$N_{ox}^{max}=DC^*/R$$

式中 C^*——饱和氧浓度；

R——细胞培养层的厚度。

9. 如下图所示，假设某个神经细胞的生长锥（growth cone）受到某个靶细胞分泌的趋化因子的作用。靶细胞的分泌速率为 P_r，趋化因子在三维空间中扩散的扩散系数为 D，那么，球形扩散源的控制方程是：

$$\frac{\partial C}{\partial t}=D\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial C}{\partial r}\right)$$

(a) 该系统的边界条件是什么？

(b) 请推导稳态时趋化因子浓度随距离 r 变化的函数。

研究结果表明，当趋化因子在生长锥上产生的浓度差大于 2%时，生长锥就会感应到靶细胞的作用。生长锥上能生出丝状伪足，并且放射状伸展出去，以增强其感受化学因子的能力。设 β_1 表示丝状伪足与中轴线 R 之间的夹角，β_2 表示反向生长的丝状伪足与中轴线 R 之间的夹角。除了图中 α 角表示的神经轴突的连接区域以外，丝状伪足可以从生长锥表面任何其他部位径向生长。

(c) β_1 和 β_2 有何限制？

(d) 请推导 r_1 和 r_2 随 β_1 变化的函数。

(e) 请计算生长锥有效直径上（即 2（$R_{gc}+l$））的百分比浓度差 $\Delta C=\frac{|C_1-C_2|}{C_0}$，它是 r_1 和 r_2 的函数，也就是 β_1 的函数。其中，C_0、C_1 和 C_2 分别表示距离 R、r_1 和 r_2 处的浓度。

(f) β_1 取何值时 ΔC 为最大？设此时的 β_1 为 β_1^{max}。请分别作出丝状伪足长度为 0μm、1μm、5μm 和 8μm 时 ΔC 随 β_1 变化的曲线，并分析结果。

(g) 生长锥能够感应的最小化学因子浓度差为 2%。如果靶细胞与生长锥相距 500μm，请计算 β_1^{max} 下 ΔC 随丝状伪足长度变化的函数曲线，并分析丝状伪足如何增强生长锥的化学因子感应能力。

有关计算公式和参数值如下：

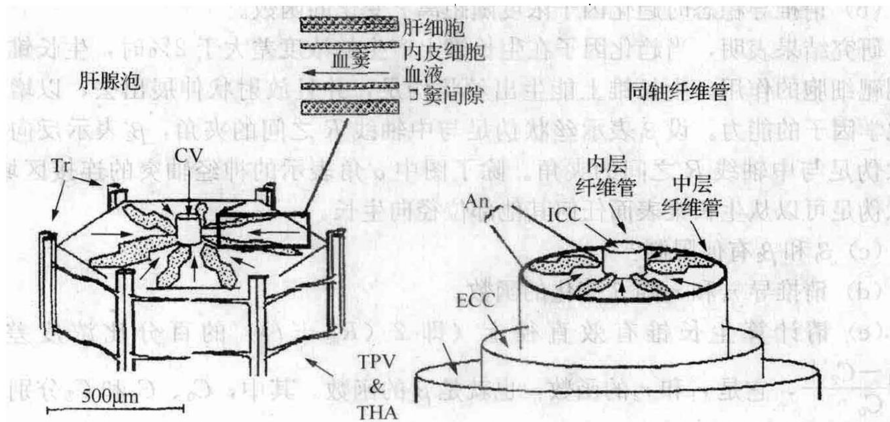
$$C_0=\frac{PrR_t^2}{Dr_o}\qquad Pr=\frac{S_r}{4\pi R_t^2}$$

$D=10^{-6}\,cm^2/s$，$S_r=5000$ 个分子/细胞/s，$R_0=R-R_{gc}$，$R_t=20\mu m$，$R_{gc}=2.5\mu m$，$\alpha=30°$。

10. 下图所示是一种模拟肝腺泡的同轴型生物反应器。

(a) 如图中所示，根据已知的肝腺泡尺寸，两个同轴纤维管之间的距离应该为多少?

(b) 根据你所了解的生理学知识，请说明建立细胞生长的微环境需要使用哪种细胞以及哪些可溶性和不溶性因子，而建立细胞分化的微环境又需要使用哪种细胞以及哪些可溶性和不溶性因子。



11. 利用房室模型计算某个特定单细胞生成的成熟细胞数量。假设细胞倍增时间（即 $\ln2/\mu$）为 24h，成熟细胞的半衰期（即 $\ln2/k_d$）为 8h；并且，自我更新只发生在第一个房室中，且更新概率为 0.5。模型一共有 10 个房室。请计算细胞数量随时间变化的函数曲线，其初始条件为：

$$X_1(0)=1,\text{其他 } X_i(0)=0(i=2,3,\cdots,10)$$

将所有 X_i 曲线画在同一张图上，并根据计算结果分析细胞移植的意义。

12. 带反馈调控的细胞分化动力学房室模型。

(a) 假设分化过程共包含 6 个阶段，即 $N=6$，其中最终的细胞群体 X_6 会产生一种细胞因子 G，并且每个细胞分泌 G 的速率为 q_G。该细胞因子的半衰期 $t_{0.5}=2\text{h}$，它会影响干细胞的生长速率（即 $u_1=f[G]$），其中 $[G]$ 为生长因子 G 的浓度。请改进细胞分化的基本微分方程组，用于描述 $[G]$ 的动态变化过程。

(b) 假设 $u_1([G])=u/(1+K[G])$，其中 K 为生长因子的结合常数。请说明函数 $f[G]$ 表示的生理意义。

(c) 请分别用生长速率和 K 作为比例因子转化时间参量和浓度参量，将方程转化为无量纲形式。

(d) 请说明各个无量纲参量的含义，并计算它们的数值。

(e) 请仿真计算由一个干细胞开始的分化过程的各个数值，并考察参数变化对于仿真结果的影响。

(f) 请计算稳态时给 X_3 的数值加上 20%扰动后，各个数值的变化过程，并分析仿真结果。

推荐阅读资料

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (1994). *Molecular Biology of the Cell,* 3rd ed. Garland, New York.

Brittberg, M. (1994). Treatment of deep cartilage defects in the knee with autologous chondrocyte transplantation. *N. Engl. J. Med.* 331, 8999.

Bruder, S.P., Fink, D.J. and Caplan, A. (1994). Mesenchymal stem cells in bone development, bone repair, and skeletal regeneration therapy. *Cell. Biochem.* **56**, 283–294.

DeWitt, L. and Wechsler, L. (1988). Transcranial Doppler. *Stroke* **19**(7), 915.

Fawcett, D.W. (1986). *A Textbook of Histology,* 11th ed. Saunders, Philadelphia.

Gage, F. (1994). Neuronal stem cells: Their characterization and utilization. *Neurobiology of Aging* 15(2). S191.

Galli, R., Binda, E., Orfanelli, U., Cipelletti, B., Gritti, A., De Vitis, S., Fiocco, R., Foroni, C., Dimeco, F. and Vescovi, A. (2004). Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. *Cancer Res.* **64(19)**, 7011–7021.

Garby, L. and Meldon, J. (1977). *The Repiratory Functions of Blood.* Plenum, New York.

Granick, S. and Levere, R. (1964). Heme synthesis in erythroidal cells. In *Progress in Hematology* (C. Moore and E. Brown, Eds.). Grune & Stratton, Orlando, FL.

Gerlyng, P., Abyholm, A., Grotmol, T., Erikstein, B., Huitfeldt, H., Stokke, T. and Seglan, P. (1993). Binucleation and polyploidization patterns in developmental and regenerative rat liver growth. *Cell Proliferation* **26**, 557–565.

Harley, C.B., Futcher, A.B. and Greider, C.W. (1990). Telomeres shorten during aging of human fibroblasts. *Nature* **345**, 458–460.

Hoffbrand, A. and Pettit, J. (1988). Clinical hematology. In *Clinical Hematology* (A. Hoffbrand and J. Pettit, Eds.). Gower Medical, London.

Iredale, J.P. (1997). Tissue inhibitors of metalloproteinases in liver fibrosis. *Int. J. Biochem. & Cell Biol.* **29**, 43–54.

Jager, Ricketts, and Strandness. (1985). *Duplex scanning for the evaluation of lower limb arterial disease. Non invasive diagnostic techniques in vascular disease.* Mosby, St. Louis.

Knauer, D.J., Wiley, H.S. and Cunningham, D.D. (1984). Relationship between epidermal growth factor receptor occupancy and mitogenic response: Quantative analysis using a steady state model system. *J. Biol. Chem.* **259(9)**: 5623–5631.

Koller, M.R. and Palsson, B.O. (1993). Tissue engineering: Reconstitution of human hematopoiesis ex vivo. *Biotechnol. & Bioeng.* **42**, 909–930.

Koller, M.R., Bradley, M.S. and Palsson, B.O. (1995a). Growth factor consumption and production in perfusion cultures of human bone marrow correlates with specific cell production. *Exper. Hematol.* **23**, 1275–1283.

Koller, M.R., Manchel, I., Palsson, M.A., Maher, R.J. and Palsson, B.O. (1995b). Different measures of ex vivo hematopoiesis culture performance is optimized under vastly different conditions. *Biotechnol. Bioeng.* **50**, 505–513.

Koller, M.R., Palsson, M.A., Manchel, I. and Palsson, B.O. (1995c). Long-term culture-initiating cell expansion is dependent on frequent medium exchange combined with stromal and other accessory cell effects. *Blood* **86**, 1784–1793.

Krstic, R. (1985). English edition, General histology of the mammal. Springer Verlag, NY, page 83, plate 40. Translated for the original German edition, Krstic, R. (1978). *Die Gewebes des Menschen und der Saugertiere.* Springer Verlag, Berlin.

Lacey, P.E., (1995, July). Treating diabetes with transplanted cells. *Sci. Am.* 50–58.

Langer, R. and Vacanti, J.P. (1993). Tissue engineering. *Science* **260**, 920–926.

Lanza, R.; Langer, R., and Vacanti, J. (2000). *Principles of Tissue Engineering*. Academic Press, Inc., San Diego, CA.

Lanza, R., Jackson, R., Sullivan, A., Ringeling, J., McGrath, C., Kuhtreiber, W. and Chick, W. (1999). Xenotransplantation of cells using biodegradable microcapsules. *Transplantation* 67(8), 1105–1111.

Lauffenburger, D.A. and Linderman, J.J. (1993). *Receptors: Models for Binding, Trafficking, and Signaling*. Oxford Univ. Press, New York.

Lightfoot, E.N. (1974). *Transport Phenoment and Living Systems*. Wiley & Sons, New York.

Lodish, H., Baltimore, D., Berk, A., Zipursky, S.L., Matsudaira, P. and Darnell, J. (1995). *Molecular Cell Biology,* 3rd ed. Scientific American Books, New York.

MacDonald, J., Griffin, J., Kubota, H., Griffith, L., Fair, J. and Reid, L. (1999). *Bioartificial Livers*. Birkhauser, Boston.

McCuskey, R. (1994). *The Hepatic Microvasculature System. The Liver: Biology and Pathology.* Raven Press Ltd., New York.

Mutsaers, S.E., Bishop, J.E., McGrouther, G. and Laurent, G.J. (1997). Mechanisms of tissue repair: From wound healing to fibrosis. *Int. J. Biochem. & Cell Biol.* **29(1)**: 5–17.

Naughton, B. (1995). The importance of stromal cells. In CRC *Handbook on Biomedical Engineering* (J.D. Bronzino, Ed.). CRC, Boca Raton, FL.

Niklason, L. and Langer, R. (2002). Morphologic and mechanical characteristics of engineered bovine arteries. *J. Vascular Surg.* **33(3)**: 628–638.

Orlowski, T., Godlewska, E., Moscicka, M. and Sitarek, E. (2003). The influence of intraperitoneal transplantation of free and encapsulated Langerhans islets on the second set phenomenon. *Artificial Organs* 27(12), 1062–1067.

Peng, C. and Palsson, B.O. (1996a). Cell growth and differentiation on feeder layers is predicted to be influenced by bioreactor geometry. *Biotechnol. & Bioeng.* **50**, 479–492.

Peng, C.A. and Palsson, B.O. (1996b). Determination of specific oxygen uptake rates in human hematopoietic cultures and implications for bioreactor design. *Ann. Biomed. Eng.* **24**, 373–381.

Rhodin, J.A.G. (1974). *Histology: Text and Atlas*. Oxford University Press, Oxford.

Rogers, C.E., Bradley, M.S., Palsson, B.O. and Koller, M.R. (1995). Flow cytometric analysis of human bone marrow perfusion cultures: Erythroid development and relationship with burst-forming ynits-erythroid (BFU-E). *Exp. Hematol.* **24**, 597–604.

Sigal, S. and Brill, S. (1992). Invited review: The liver as a stem cell and lineage system. *Am. J. Physiol.* **263**, 139–148.

第8章　生物医学仪器

John Enderle 博士[⊖]

本章目录

⊖ Susan Blanchard，Amanda Marley 和 H. Troy Nagle 也参与了本章部分内容的撰写。——原文注

本章主要内容：

- 基本生物医学仪器系统的组成；
- 利用节点电压法分析线性电路；
- 利用戴维南（Thevenin）等效电路简化复杂电路；
- 包含电阻、电容和电感的任意阶电路的求解方法；
- 运算放大器电路的分析方法；
- 在相量域中计算正弦波输入的稳态响应；
- 模拟滤波器设计的基本原理；
- 低通、高通和带通滤波器的设计；
- 生物医学仪器系统中的各种噪声。

8.1 绪论

本章介绍其他章节要用到的生物医学仪器和电子电路理论的基本内容。正如第9章所述，很多生物医学仪器都利用传感器将生物体所产生的各种信号转化为电信号。本章的目的是介绍仪器中应用的电路理论和方法。首先讲述电路理论所涉及的电荷、电流、电压、功率和能量等变量的概念；然后介绍基尔霍夫电流定律和电压定律，以及电阻的概念、电阻电路的简化方法、电压和电流的计算等；再讲解电路分析方法电感和电容以及利用微分方程求解电路的方法。最后，介绍运算放大器和时变信号。

20世纪前，普通市民很少接触医学，当时的医学资源主要用于教学，是医生的秘密。最早的医学革命源于科学技术应用的几项重大发展。19世纪初，医学诊断还只是基于体格检查，治疗只是处理器官和组织的结构性病变。直到19世纪末，医学界才开始根据实验室的化验结果来进行医学诊断，并针对发病原因进行治疗。整个20世纪医疗领域的应用技术迅速发展，医院变成了研究和技术开发的机构。化学、物理、机械工程以及电子工程等领域的从业人员开始与医学界相结合，于是，产生了生物医学工程这一职业。如图8.1所示，这使得20世纪医学科技的发展超过了之前其他历史阶段的总和。

20世纪电子技术的发展对于新型医学技术的发明产生了巨大的影响。Richard Caton和Augustus Desire等人证明了人的大脑和心脏的工作过程都与生物电密切相关。1903年，William Einthoven发明了第一台弦线式检流计，并用它测量了生物电。

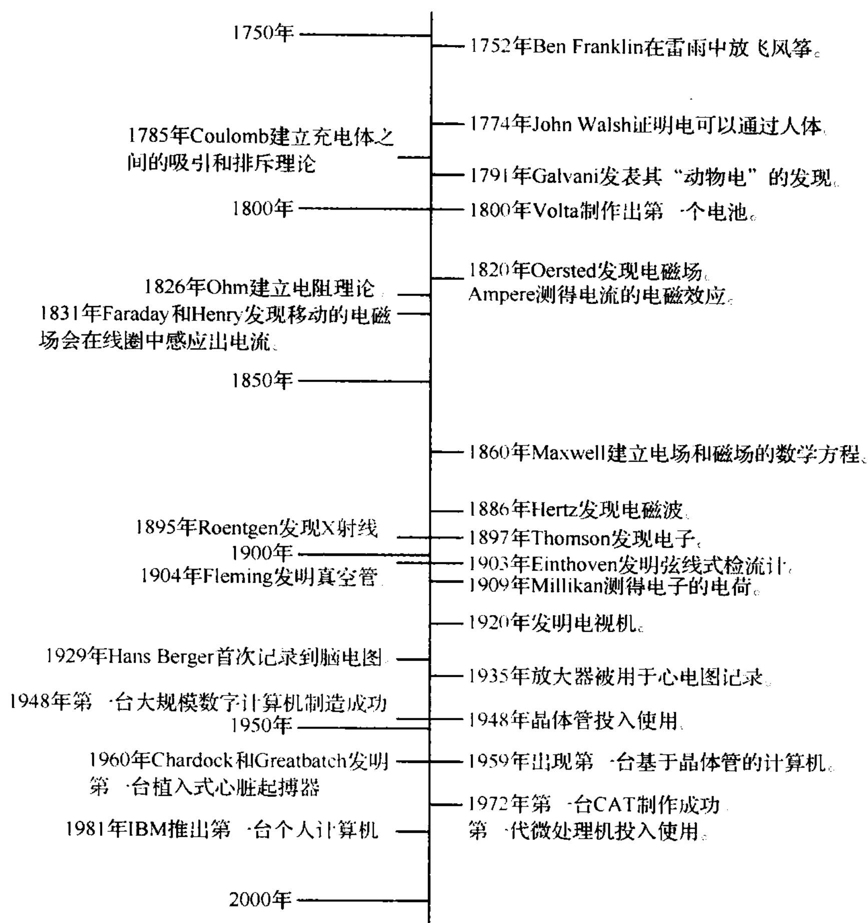


图 8.1 现代医学仪器发展过程中主要的发明和发现时间表

他在人体皮肤上放了两个电极，并将电极的输出连接到一根银导线上，这根银导线悬空穿过一块大永磁铁上所打的孔，悬挂的导线会随着人体心脏的跳动有节律地移动。通过银导线反射一小束光线的方法，Einthoven 在走动的感光筒纸上记录到了导线的移动轨迹。于是，这种弦线式检流计的发明就构成了最早的心电图（ECG）机。如今，ECG 已经是临床的常规检查方法，用于检测和记录病人心脏的电活动，并与正常心脏进行比较。

1929 年，Hans Berger 发明了第一台脑电图（EEG）机，用于检测和记录大脑的电活动。1935 年，人们利用电子放大器证实了大脑皮层的电活动具有特定的节律。到了 1960 年，William Chardack 和 Wilson Greatbatch 应用电子放大器，发明了第一台植入式心脏起搏器。这些只是少数几个例子，电子技术的应用给医学带来的重大发展数不胜数。

还有很多其他医学技术的进展源于基础物理学和应用物理学的研究成果。例如，

1895年，W. K. Roentgen发现的X射线可以用于摄取人体内部结构的图像。从此产生了医学领域最重要的技术之一——X射线摄影机，它是历史上第一种医学成像设备。本书第15章将详细介绍这种放射成像技术。

计算机的发明给医学带来了又一项重要的技术。利用计算机可以快速完成复杂的计算分析、增强仪器的功能。例如，医学领域最早出现的计算机化仪器之一就是多通道自动生化分析仪，它可以存储大量临床化验数据，提高计算速度和精度。

最早的大型计算机医学设备是1972年发明的X射线计算机轴向断层扫描机(Computerized Axial Tomography，CAT)。CAT可以建立并显示人体各个断层平面上所有内部结构的图像。这种新型的图像精度很高，可用于肿瘤、出血部位等体内病变的无创式诊断，有关这方面的详细内容将在第15章介绍。

目前，通过计算机将医学信息从某处传到另一处的远程医疗也正在迅速发展，它使位于不同地方的病人都可以享受医疗保健服务。大医院的专家利用此项技术可以接收偏远地区病人的资料，并将病人的医疗方案发送回去。

如今，各种各样的医学仪器和设备有很多。有些用于监测病人的身体状况、采集诊断信息，如心电图机和脑电图机等；还有一些则用于调控生理功能，如心脏起搏器和呼吸机等。心脏起搏器等装置是植入式的，许多其他仪器则是无创的。本章将着重介绍用于采集和处理生理数据的医学仪器的共同特性。

8.2 基本生物医学仪器系统

如图8.2所示，仪器系统所测量的参量、特性和状态等都称为被测量。被测量可以是肌肉或大脑等产生的生物电信号，也可以是由化学量或者机械量等转换而来的电信号。就像第9章所述，传感器可以将各种其他物理量转换成电信号，并且传感器输出的电信号是模拟信号，也就是连续信号，可以传输给模拟处理器和模/数转换器。再由这些器件对信号进行放大、滤波并转换成数字信号。本章后面将介绍模拟信号的处理方法，如心电图信号的放大和滤波等。模拟信号经过数字化之后，才能转换成可以由数字计算机存储和处理的信号。本书第10章将详细介绍数字信号处理的各种方法。

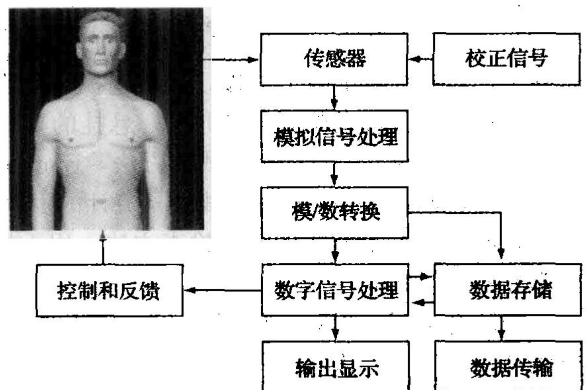


图8.2 带传感器的基本生物医学仪器系统
(它具有数据采集、存储和显示功能，
还有控制和反馈功能)

仪器系统还包括输出显示装置，它使操作人员可以观察到易于理解的检测信号。显示有很多种不

同的方式：如数据或图形形式、离散或连续形式、永久性的或暂时性的等。一般输出装置都显示视觉信息，但有些也提供听觉信息，如代表每次心跳的蜂鸣声等。

除了显示数据之外，许多仪器还具有存储数据的功能。有些仪器只是暂时存储数据，以便于进一步处理，或者便于用户检查数据。但有些仪器则永久保存数据，便于事后用于各种不同的处理。例如，Holter 监护仪可以采集并存储 24h 的心电图数据，事后再计算心率等重要的特征参数用于诊断。

随着电话的发明，以及如今互联网的发明，病人家中等其他地方采集的信号也可以方便地传输到远方的仪器设备上进行处理和存储。这样，如果病人在家里产生了心率失常，就可以快速收到诊断反馈意见。这也使得处于偏远乡村的医疗机构可以将病人图像传输给地区医院，让那里的专家帮助乡村普通医务人员作出更准确的诊断。

仪器系统还有另外两个重要的部分。一个是校正信号，它是具有已知幅值和频率的标准信号。将其加到仪器的输入端，就可以调整系统的各个部件，使输出和输入之间保持正确的测量关系。如果没有校正信号，就不能获得有意义的被测参量的输出结果。

另一个重要部件是反馈部件，它并不是所有仪器都具备的。刺激心脏的起搏器或者给肺通气的呼吸机等仪器具有反馈部件。其中，有些反馈部件可以根据所采集的生理数据，按需提供触发心跳或者呼吸等事件的信号；有些反馈则作为生物反馈的一部分，如血压测量。病人知道这种生理参数的测量结果之后，会有意识地控制和调节生理状态，这就是生物反馈。

8.3 电荷、电流、电压、电功率和电能量

8.3.1 电荷

电荷有正、负两种，质子和电子分别带有正电荷和负电荷。电荷的单位是库仑(C)，电子所带的负电荷 q_e 是最小的电荷量，即

$$q_e = -1.602 \times 10^{-19}\,\mathrm{C}$$

符号 $q(t)$ 用于表示随时间变化的电荷，Q 则表示电荷常量。质子所带的电荷与电子所带电荷极性相反。

8.3.2 电流

电流 $i(t)$ 定义为单位时间内通过某个点或者某个截面的电荷量。电流的单位是安培（A）。根据定义，1A 等于 1C/s。即

$$i(t) = \frac{\mathrm{d}q}{\mathrm{d}t} \tag{8.1}$$

并且

$$q(t)=\int_{t_0}^{t} i(\lambda)\mathrm{d}\lambda+q(t_0) \tag{8.2}$$

如图 8.3 的电路所示，式（8.1）定义的电流还具有方向。如果正电荷沿着箭头所示方向运动，负电荷沿反方向运动，那么，此电流为正。由于这两种电荷反方向流动所产生的效果相同，因此，无需确定电流是由哪种电荷形成的。实际上，电子电路中的电流是由金属导体中的电子形成的。

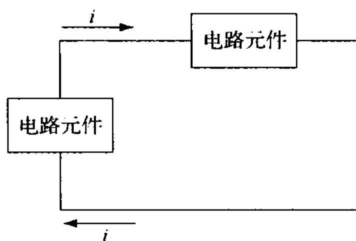


图 8.3 电流在闭合电路中流动的示意图

由式（8.1）可知，电流是时间的函数。例如，图 8.4 所示的电流曲线在 0～1.5s 时间段上，电流为正，它从电路的端口 1 流入；而在 1.5～3s 时间段上，电流为负，从电路的端口 2 流入。恒定不变的电流称为直流电流（Constant Current，DC)，用大写字母 I 表示，其值不随时间变化。随时间变化的电流则用小写字母 i（t）或 i表示。

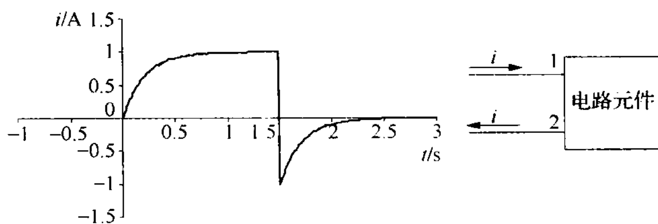


图 8.4 左图：电流波形示例；右图：某电路元件，电流从端口 1 流入，从端口 2 流出

注：电阻、电容、电感等无源电路元件都具有两个端口，并且它们的电压一电流关系已知。

基尔霍夫电流定律

如图 8.3 所示，闭合电路中才有电流流动。电路元件不能吸收净电荷，因此，电路中流动的电流没有损耗，电荷是守恒的，流入电路一端的电流必须在另一端流出。由于电荷不能产生或消失，必定守恒，因此，任何节点上的电流之和必为零，没有净电荷聚集。这里，节点是指两个或两个以上电路元件的连接点。上述原理称为基尔霍夫电流定律（Kirchhoff's Current Law，KCL)，可以表示为

$$\sum_{n=1}^{N} i_n(t)=0 \tag{8.3}$$

式中 N——流出节点的电流数目。

例如，对于图 8.5 所示的电路，应用式（8.3），流出节点的电流之和为

$$-i_1-i_2+i_4+i_3=0$$

该式也可以写为流入节点的电流之和，即

$$i_1+i_2-i_4-i_3=0$$

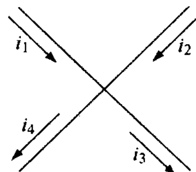


图 8.5 包含 4 路电流的某个电路节点

注意，KCL 必须包括所有流出和流入节点的电流。

电路可以用节点、支路、路径、回路和网孔等术语来描述，它们定义如下：

1）节点（node）：指两个或两个以上电路元件相互连接的共同点。

2）支路（branch）：电路中通过同一电流的各个分支，由一个或者一组电路元件构成。支路将各个电路节点连接起来。

3）路径（path）：任意两节点之间的一条无重复通路，它由支路构成。

4）回路（closed path）：起点和终点在同一个节点上的闭合路径。

5）网孔（mesh）：内部不包含其他任何回路的闭合路径。

6）基本节点（Essential node）：3 个或者 3 个以上电路元件的连接点。

7）基本支路（Essential branch）：连接两个基本节点的支路。

例如，图 8.6 所示的电路有 A、B、C、D、E 5 个节点，它们都是基本节点。对于这 5 个节点分别应用 KCL 可得

节点 A：$-i_1+i_2-i_3=0$

节点 B：$i_3+i_4+i_5-i_6=0$

节点 C：$i_1-i_4-i_8=0$

节点 D：$-i_7-i_5+i_8=0$

节点 E：$-i_2+i_6+i_7=0$

KCL 还可以推广到包含了部分电路的封闭面。当然，封闭面不能与电路元件相交。例如，对图 8.6 中点划线所示的封闭面应用 KCL，可得

$$-i_1+i_4+i_5+i_7=0$$

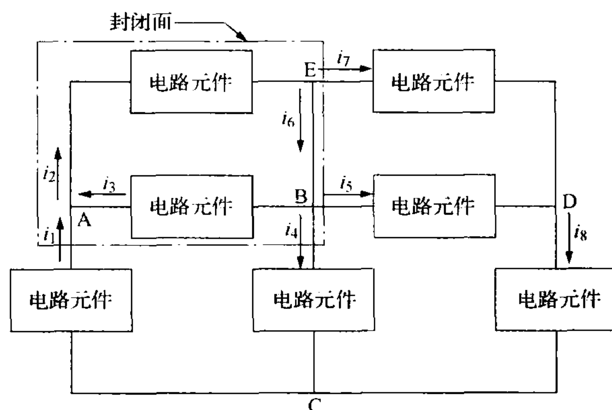


图 8.6　带封闭面的电路

8.3.3　电压

将单位正电荷从一点移至另一点时电场力所做的功就是电压。例如，图 8.7 中 A 和 B 两点之间的电压为

$$v=\frac{dw}{dq} \tag{8.4}$$

电压的单位是伏特（V）。不随时间变化的直流电压（DC）用 V 表示，随时间变化的电压用 v（t）或 v 表示。图 8.7 中 A 和 B 两点之间的电压 v 就是将单位电荷从 A 点移动到 B 点所需做的功。

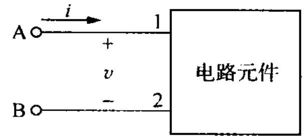


图 8.7 电压的定义

基尔霍夫电压定律

基尔霍夫电压定律（Kirchhoff's Voltage Law，KVL）表明任一闭合回路中所有电压之和为零，即

$$\sum_{n=1}^{N} v_n(t) = 0 \tag{8.5}$$

式中 N——沿闭合回路的电压数；

$v_n(t)$——各个电压值。各电压的正负号是绕回路行走时首先遇到的那个极性符号。

例如，设图 8.8 所示电路中的各个电路元件两端的电压为 v_n 并带有极性，该电路共有 CP1、CP2 和 CP3 3 个回路。对这 3 个回路分别应用 KVL，可得

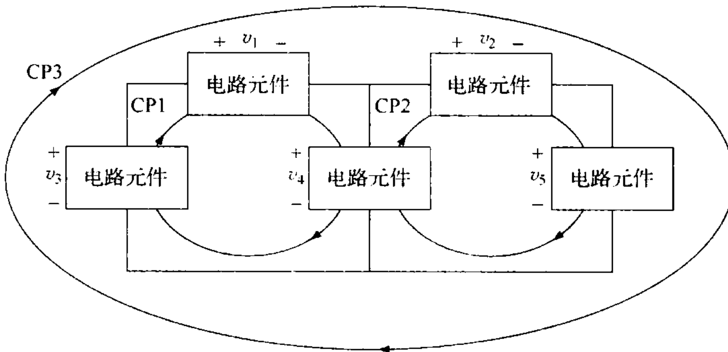


图 8.8 基尔霍夫电压定律的电路示意图（其中 CP1、CP2 和 CP3 表示回路）

$$\text{CP1}: -v_3 + v_1 + v_4 = 0$$
$$\text{CP2}: -v_4 + v_2 + v_5 = 0$$
$$\text{CP3}: -v_3 + v_1 + v_2 + v_5 = 0$$

将上述两个基尔霍夫定律用于电路分析，可以求取出电路中未知的电压和电流。首先，对各个未知量建立电路方程，组成一个方程组；再将其表示成矩阵形式；然后，就可以用本书附录所述的方法，编写 MATLAB 程序求解方程组的未知量。本章下面所述的许多例题都会用到这种方法。

8.3.4 电功率和电能量

电功率是电能消耗的速率，即

$$p = \frac{dw}{dt} = \frac{dw}{dq}\frac{dq}{dt} = vi \tag{8.6}$$

式中 p——电功率，单位是 W；

w——电能量，单位是 J。

可见，功率等于电路元件两端的电压与通过该电路元件的电流的乘积。按照惯例，功率为正时，表示电路元件吸收或损耗的功率；功率为负时，表示电路元件（即电源）释放或产生的功率。

图 8.9 显示了电路元件上电流和电压的 4 种可能的组合情况。根据规定，如果如图 8.9A 的箭头和极性所示，i 和 v 都为正，则能量被电路吸收，要么发热损耗掉，要么被存储起来；如果如图 8.9 中的 B 和 C 所示，把电流的方向反过来，或者把电压的极性反过来，则电路就是提供能量；如果如图 8.9 中的 D 所示，把电流的方向和电压的极性都一起反过来，则能量还是被电路吸收。

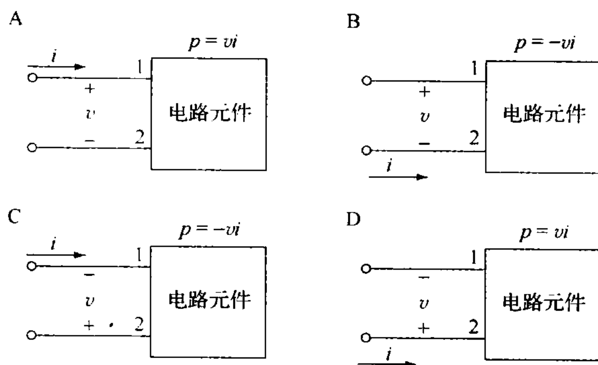


图 8.9 电流和电压的 4 种极性组合情况

注：A 和 D 产生正功率，电路元件消耗能量；B 和 C 产生负功率，电路元件提供能量。

无源电路元件的功率总是正值或者零，其能量要么被电阻的发热消耗、要么被电容存储在电场中、要么被电感存储于磁场中。有源电路元件能够产生能量，因此其功率为负值。

电能量的计算公式为

$$w(t) = \int_{-\infty}^{t} p\mathrm{d}t \tag{8.7}$$

8.3.5 电源

电源是一种双端口装置，可以给电路提供电能。它没有固定不变的电压—电流关系，即使已知电压和电流两个变量中的某一个，如果不知道电路其余部分的信息，也不能求得另一个变量。电源可分为独立电源和受控电源两类。独立电源（independent source）不管电路的其余部分如何，都能够保持一定的输出电压或者输出电流。其中，理想电压源无论通过的电流有多大都可以在输出端产生固定的标称电压。图 8.10a 和 b 所示是理想电压源的通用符号。理想电流源则可以给外接电路提供标称电流，图 8.10c 所示是其表示符号。理想电流源所产生的电压则由电路其余部分决定。

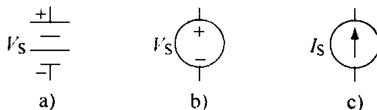


图 8.10 独立电源的表示符号

a）电池 b）理想电压源，其电压 V_S 可以是直流电压（即电池），也可以是随时间变化的电压 c）理想电流源 I_S

图 8.11 所示是受控电压源和受控电流源的表示符号。受控电源（controlled source）也称非独立电源（dependent source），其输出值是电路中其他电压或者电流的已知函数。我们用菱形符号表示受控电源。受控电压源所产生的电流和受控电流源所产生的电压都由电路其余部分的元件决定。受控电源在电子学中具有非常重要的作用。本章后面所述的运算放大器就是基于受控电压源的原理。

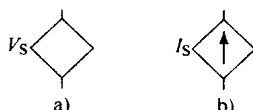


图 8.11 受控电源的符号表示

a）受控电压源，其电压 V_S 是电路中其他某个电压或者电流的已知函数

b）受控电流源，其电流 I_S 是电路中其他某个电压或者电流的已知函数

8.4 电阻

8.4.1 电阻的定义

电阻是阻碍电流流过的一种电路元件，用符号⋀⋀⋀表示。电阻可以用各种不同的材料制作，其阻抗值 R 决定它阻止电流的能力。电阻的单位是欧姆（Ω），1Ω=1V/A。理论上，连接电路元件的导线的电阻为零；而断开的电路元件两端之间的电阻为无穷大。理想电阻遵循欧姆定律，该定律描述了电压与电流之间的线性关系，其斜率就等于电阻值。

根据电流的方向和电压的极性，欧姆定律有两种写法：图 8.12a 的欧姆定律为

$$v=iR \tag{8.8}$$

而图 8.12b 的欧姆定律为

$$v=-iR \tag{8.9}$$

本书使用图 8.12a 所示的方式表示电阻上的电压降。可见，电阻两端的电压等于流过其中的电流 i 与阻抗 R 的乘积。但是，在高电压和大电流的情况下，这种线性关系可能不成立。某些导电材料只在非常小的电流和电压范围内才具备线性特性。许多生理模型也是如此，只有在一定的数值范围内才是线性的，超出这个范围，模型就是非线性的。图 8.13a 所示的短路电路定义为 $R=0$，其电压降为 0V；而图 8.13b 所示的开路电路则定义为 $R=\infty$，通过的电流为 0A。

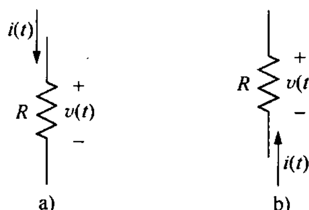


图 8.12 具有电阻 R 的理想电阻

每种材料都具有体现其阻抗的电阻率 ρ（resistivity）。电导率 σ（conductivity）为电阻率的倒数，电导 G 则是电阻的倒数。电导的单位是西门子（S），即 A/V。因此，欧姆定律也可以写成电导的形式，即

$$i=Gv \tag{8.10}$$

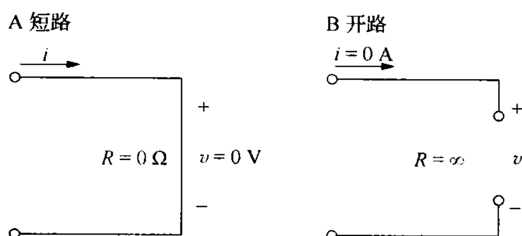
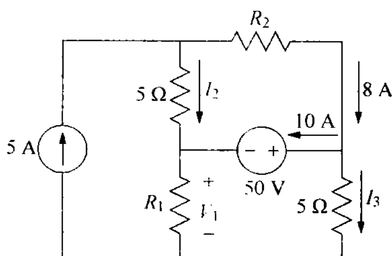


图 8.13 短路电路和开路电路

例 8.1 请计算以下电路中的 I_2、I_3 和 V_1。



解：首先，对电路左上方的节点应用 KCL，求 I_2，即

$$-5+I_2+8=0$$

则

$$I_2=-3\text{A}$$

然后，对电路右边的节点应用 KCL，求 I_3，即

$$10+I_3-8=0$$

则

$$I_3=-2\text{A}$$

最后，对电路右下方回路应用 KVL 和欧姆定律，求 V_1，即

$$-V_1-50+5I_3=0$$

$$V_1=-50+5\times(-2)=-60\text{V}$$

■

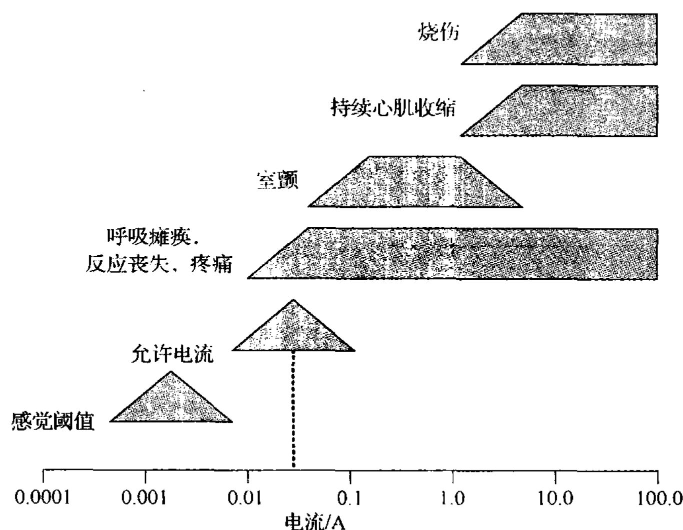
8.4.2 电阻的功率

将式（8.8）或者式（8.9）代入式（8.6）可以得到电阻上消耗的功率的计算公式，即

$$p=vi=\frac{v^2}{R}=i^2R \tag{8.11}$$

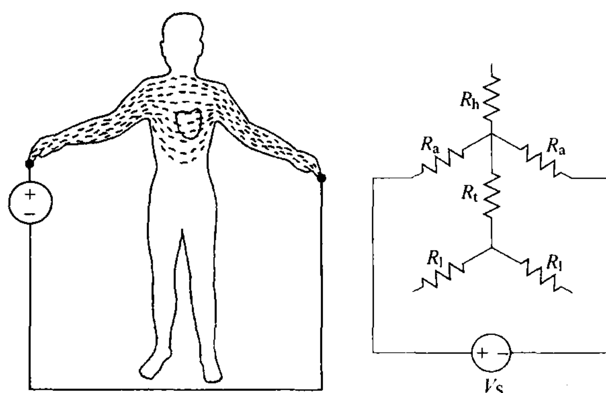
该功率以发热的形式耗散出去。由此计算公式可见，电阻所消耗的功率与电压极性和电流方向都无关，电阻的功率总是正的，任何其他无源元件也都是如此。

例 8.2

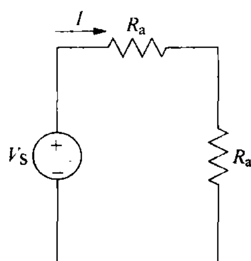


医院临床环境下的电气安全极为重要。如上图所示，如果流过人体的电流比较大，就会导致严重的伤害。例如，50mA 的电流（图中虚线所示）就可能引起室颤等症状。

下面这张图的左边显示了人体两手之间受到电击时人体中的电流分布。右边显示了其粗略的电路模型，包括阻抗各为 R_a 的两条手臂，阻抗各为 R_l 的两条腿，阻抗为 R_t 的躯干，以及阻抗为 R_h 的头部。



可见，电源与两条手臂串联并形成了惟一的闭合回路，因此，该人体电路模型可以简化为



如果 $R_a=400\Omega$，$V_S=120V$，请计算电流 I 的值。

解：利用欧姆定律，有

$$I=\frac{V_S}{R_a+R_a}=\frac{120}{800}=0.15A$$

I 就是流过心脏的电流，该电流值会导致室颤。■

8.4.3 等效电阻

复杂电路经常可以简化为简单的等效电路（equivalent circuit）。如果两个电路的电压和电流没有区别，也就是两个电路的性能一样，那么，就可以认为两个电路等效。例如，图 8.14 的电路 A 和电路 B 都由电阻组成，外加同样的直流电压 V_S，如果 $I_A=I_B$，那么两个电路就等效。根据欧姆定律，两个电路的阻抗均为

$$R_{EQ}=\frac{V_S}{I_A}=\frac{V_S}{I_B} \tag{8.12}$$

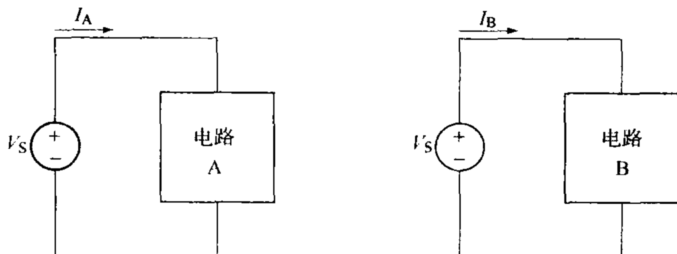


图 8.14　两个电路

由此可知，任何电阻组成的电路都可以用图 8.15 所示的等效电路来代替。在下面讲到戴维南等效电路（Thevenin equivalent circuit）时，还可以将这种方法推广到电源和电阻组成的任何电路。

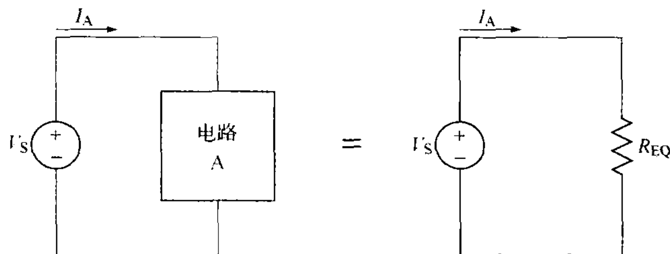


图 8.15　等效电路

8.4.4 电阻的串联和并联

串联电阻

如果两个电阻上流过的是同一个电流，那么这两个电阻就是串联；如果这两个

电阻再与第三个电阻相连，并且通过 3 个电阻的电流都相同，那么这 3 个电阻是串联。一般，如果同一个电流流过 N 个电阻，那么这 N 个电阻就是串联的。例如，图 8.16 是 3 个电阻串联。利用 KVL 可以求其等效电阻，即

$$-V_S+IR_1+IR_2+IR_3=0$$

等效电阻 R_{EQ} 即为

$$R_{EQ}=\frac{V_S}{I}=R_1+R_2+R_3$$

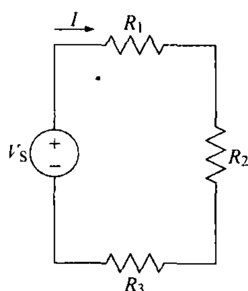


图 8.16　串联电路

一般，N 个串联电阻的等效电阻 R_{EQ} 为

$$R_{EQ}=\sum_{i=1}^{N}R_i \tag{8.13}$$

并联电阻

如果两个或两个以上电阻中每个电阻两端的电压都是同一个，那么这些电阻就是并联。如图 8.17 所示是 3 个电阻并联。我们用符号 ∥ 表示并联电阻。于是，图 8.17 可以表示为 $R_{EQ}=R_1 \parallel R_2 \parallel R_3$。利用 KCL 可以求出其等效电阻，即

$$-I+\frac{V_S}{R_1}+\frac{V_S}{R_2}+\frac{V_S}{R_3}=0$$

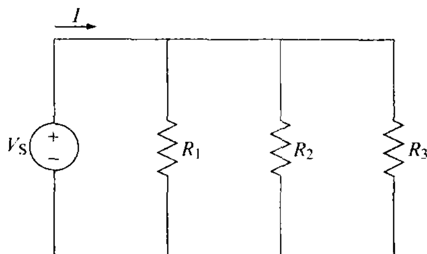


图 8.17　并联电路

等效电阻 R_{EQ} 即为

$$R_{EQ}=\frac{V_S}{I}=\frac{1}{\frac{1}{R_1}+\frac{1}{R_2}+\frac{1}{R_3}}$$

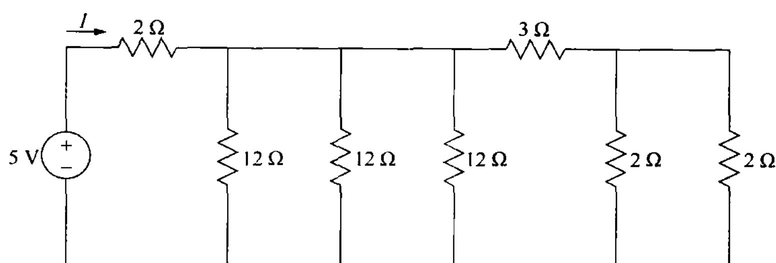
一般，N 个并联电阻的等效电阻 R_{EQ} 为

$$R_{EQ}=\frac{1}{\frac{1}{R_1}+\frac{1}{R_2}+\cdots+\frac{1}{R_N}} \tag{8.14}$$

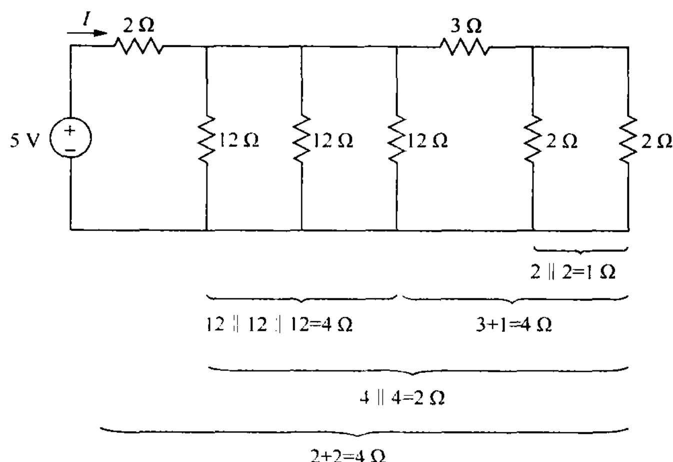
对于两个并联电阻，上式成为

$$R_{EQ}=R_1 \parallel R_2=\frac{R_1R_2}{R_1+R_2} \tag{8.15}$$

例 8.3　请计算以下电路的等效电阻 R_{EQ} 和电源所提供的功率。



解：从右到左分析电阻的并联和串联关系，可以求得 R_{EQ}。首先是两个 2Ω 电阻并联，再与一个 3Ω 电阻串联；然后，这组电阻与 3 个 12Ω 电阻并联；最后，再与一个 2Ω 电阻串联。下图表示了这些组合关系及其计算结果。



$$R_{EQ}=2\Omega+((12\Omega \parallel 12\Omega \parallel 12\Omega) \parallel (3\Omega+(2\Omega \parallel 2\Omega)))$$

$$=2+\left(\left(\frac{1}{\frac{1}{12}+\frac{1}{12}+\frac{1}{12}}\right) \parallel \left(3+\frac{1}{\frac{1}{2}+\frac{1}{2}}\right)\right)$$

$$=2+(4 \parallel (3+1))=2+2=4\Omega$$

于是，电源的输出电流为

$$I=\frac{5}{R_{EQ}}=\frac{5}{4}=1.25\text{A}$$

功率为

$$p=5\times I=6.25\text{W}$$

■

8.4.5　分压器和分流器

将上述等效电阻 $R_{EQ}=\frac{V}{I}$ 的概念加以推广，就可以快速计算串联电阻电路中的电压或者并联电阻电路中的电流。

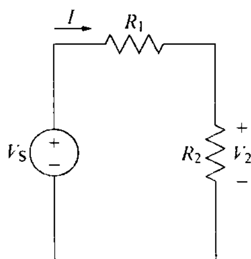


图 8.18　分压器电路

分压器原理

根据分压器原理可以方便地计算串联电路中指定电阻两端的电压。例如，要计算如图 8.18 所示的串联电路中的 V_2，电路的等效电阻为 $R_{EQ}=R_1+R_2$，因此

$$I=\frac{V_S}{R_{EQ}}=\frac{V_S}{R_1+R_2}$$

于是

$$V_2=IR_2=V_S\frac{R_2}{R_1+R_2}$$

同理，可以求得 V_1 为

$$V_1=V_S\frac{R_1}{R_1+R_2}$$

一般，如果电路中含有 N 个串联电阻，则由分压器规则可求得其中任意一个电阻 R_i 两端的电压为

$$V_i=V_S\frac{R_i}{R_1+R_2+\cdots R_N} \tag{8.16}$$

分流器原理

根据分流器原理可以方便地计算并联电路中任意一个电阻上流过的电流。例如，要计算如图 8.19 所示的并联电路中的 I_2，电路的等效电阻为 $R_{EQ}=\frac{R_1R_2}{R_1+R_2}$，因此

$$I_2=\frac{V_S}{R_2}$$

并且

$$V_S=I\frac{R_1R_2}{R_1+R_2}$$

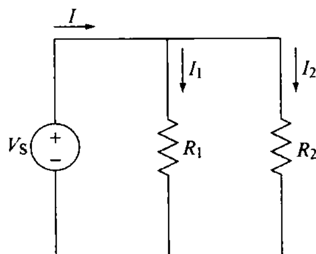


图 8.19　分流器电路

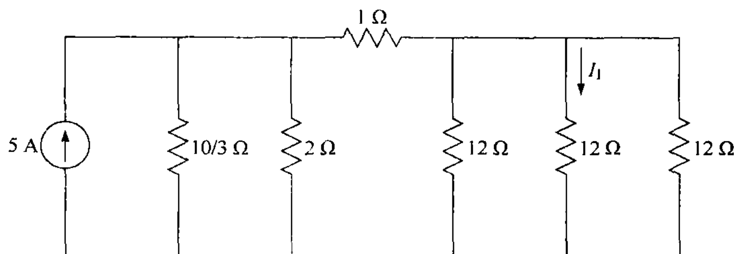
将 V_S代换掉以后可得

$$I_2=I\frac{\frac{1}{R_2}}{\frac{1}{R_1}+\frac{1}{R_2}}$$

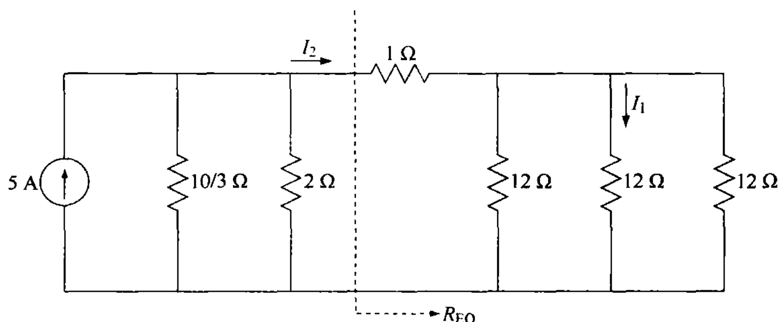
一般，如果电路中含有 N 个并联电阻，则由分流器原理可以求得其中任意一个电阻 R_i 的电流为

$$I_i = I\frac{\frac{1}{R_i}}{\frac{1}{R_1}+\frac{1}{R_2}+\cdots+\frac{1}{R_N}} \tag{8.17}$$

例 8.4　请计算以下电路中的电流 I_1。



解：该问题需要分两步求解，将电路重新画成如下形式就很清楚了。先求电流 I_2，再求电流 I_1。



首先计算出图中所示的 R_{EQ}，将其代入电路之后，电路就可以简化成 3 个电阻的并联，从而求得 I_2。等效电阻 R_{EQ} 为

$$R_{EQ}=1+(12 \parallel 12 \parallel 12)=1+\frac{1}{\frac{1}{12}+\frac{1}{12}+\frac{1}{12}}=5\Omega$$

将分流器原理用于 3 个并联电阻 $\frac{10}{3} \parallel 2 \parallel R_{EQ}$，可得

$$I_2=5\left(\frac{\frac{1}{5}}{\frac{3}{10}+\frac{1}{2}+\frac{1}{5}}\right)=1\text{A}$$

该电流 I_2 流过 1Ω 电阻，然后分成三等份（即 1/3A），流过每个 12Ω 电阻。因此，应用分流器原理就可以求得电流 I_1，即

$$I_1=I_2\left(\frac{\frac{1}{12}}{\frac{1}{12}+\frac{1}{12}+\frac{1}{12}}\right)=\frac{\frac{1}{12}}{\frac{1}{12}+\frac{1}{12}+\frac{1}{12}}=\frac{1}{3}\text{A}$$ ■

8.5 线性电路网络分析法

至此，我们已经学习的电路求解方法有：欧姆定律、基尔霍夫定律、等效电路简化以及分压器和分流器。这些方法可以用于任何电路问题的求解。不过，电路越复杂，求解就越困难。本节将介绍的节点电压法（node-voltage method）提供了一种方便的、系统性较强的电路问题求解的一般方法。该方法将支路电流用一个或多个节点电压来表示，并且对电路中每个节点应用 KCL，建立一组方程来求解电路。它是一种能够有效可靠地求解电路问题的系统方法，所产生的联立方程数目最少，可以减少工作量并节省时间。

节点电压法可以分为如下两个求解步骤：

1）设定一个参考节点，也就是"地"，电路中其他各个节点（称为基节点）的电压就是相对于参考节点的电压。通常取电路中与其相连支路最多的节点作为参考节点，并用符号⏚表示。

2）对电路中参考节点以外的每个节点（即 $N-1$ 个节点）应用 KCL 写出节点电流方程。

如图 8.20 所示，应用欧姆定律写出流过电阻的电流，其中所用的电压就是电阻两端相对于参考节点的两个电压之差。对于两个相邻节点，一个节点的电流向右（见图 8.20a），而另一个节点的电流则向左（见图 8.20b）。于是，图 8.20a 的电流为 $I_A=\frac{V}{R}=\frac{V_1-V_2}{R}$，而图 8.20b 的电流为 $I_B=\frac{V}{R}=\frac{V_2-V_1}{R}$。利用 KVL 很容易证明图 8.20a 中 $V=V_1-V_2$。

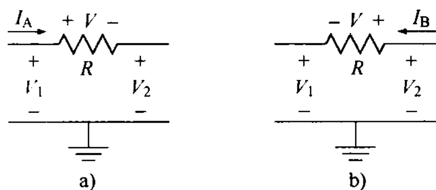
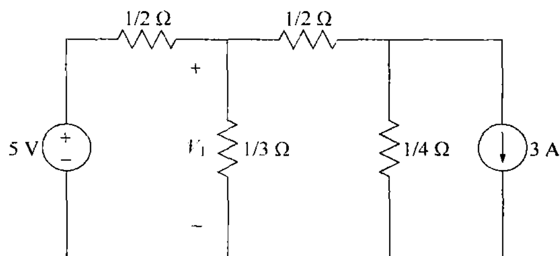


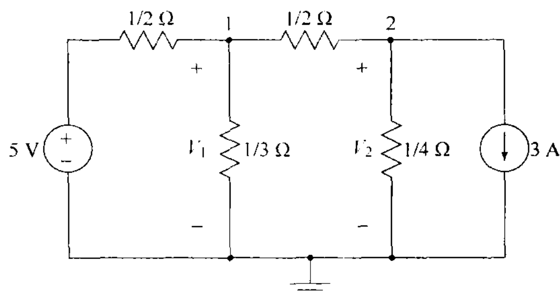
图 8.20 使用节点电压写出欧姆定律方程

如果在基节点和参考节点之间的某条支路上含有独立电压源或者受控电压源，那么，该节点的电压已知，就不用写节点方程了。这可以减少独立的节点方程数目，从而减少求解节点电压的工作量。而且，在写其他节点方程时可以直接使用独立电压源的数值。例如，假设图 8.20a 中的电压 V_2 是一个 5V 独立电压源。由于图中右边节点的电压已知，该节点就不要写节点方程了，在写左边节点的方程时，电流 I_A 就等于 $I_A=\frac{V_1-5}{R}$。下面的例 8.5 将进一步说明这种情况。

例 8.5 请使用节点电压法求出如下电路中的电压 V_1。



解： 将电路图重画如下。可见该电路有两个基节点，分别用“1”、“2”标出，并且用 V_1 和 V_2 表示这两个节点与参考节点之间的电压。图中另一个含有 5V 电源的节点因电压已知，不必写节点方程。



流出节点 1 的电流之和为

$$2(V_1-5)+3V_1+2(V_1-V_2)=0$$

可以简化为

$$7V_1-2V_2=10$$

流出节点 2 的电流之和为

$$2(V_2-V_1)+4V_2+3=0$$

可以简化为

$$-2V_1+6V_2=-3$$

将这两个节点方程写成矩阵形式，即

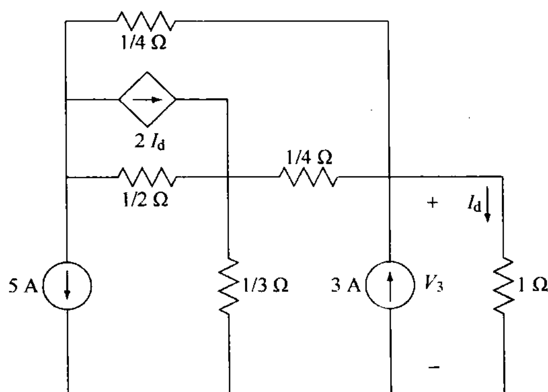
$$\begin{bmatrix} 7 & -2 \\ -2 & 6 \end{bmatrix}\begin{bmatrix} V_1 \\ V_2 \end{bmatrix}=\begin{bmatrix} 10 \\ -3 \end{bmatrix}$$

再用如下 MATLAB 程序求解该方程组，即

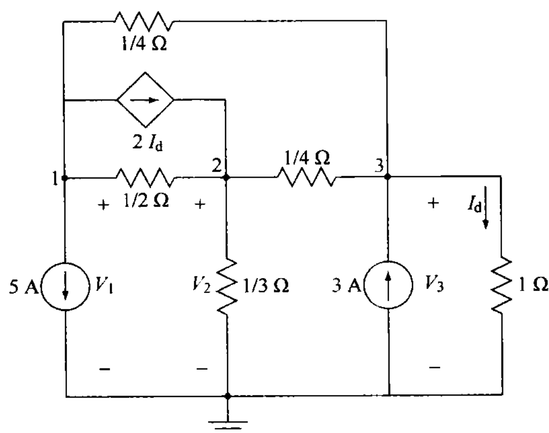
```
≫A=[7  -2; -2  6];
≫F=[10; -3];
≫V=A\F
V=
1.4211
-0.0263
```

可得 $V_1=1.4211$V。 ■

例 8.6 请使用节点电压法求出如下电路中的电压 V_3。



解： 注意，该电路有 3 个基节点和一个受控电流源。在如下重画的电路图中，分别用"1"、"2"和"3"标出了 3 个基节点，并且标出了电路底部的参考节点，以及 3 个节点电压 V_1、V_2和 V_3。



注意，根据欧姆定律，$I_d=V_3$。

流出节点 1 的电流之和为

$$5+2(V_1-V_2)+2I_d+4(V_1-V_3)=0$$

可以简化为

$$6V_1-2V_2-2V_3=-5$$

流出节点 2 的电流之和为

$$-2I_d+2(V_2-V_1)+3V_2+4(V_2-V_3)=0$$

可以简化为

$$-2V_1+9V_2-6V_3=0$$

流出节点 3 的电流之和为

$$4(V_3-V_2)-3+V_3+4(V_3-V_1)=0$$

可以简化为

$$-4V_1-4V_2+9V_3=3$$

将这 3 个节点方程写成矩阵形式，即

$$\begin{bmatrix}6 & -2 & -2\\ -2 & 9 & -6\\ -4 & -4 & 9\end{bmatrix}\begin{bmatrix}V_1\\ V_2\\ V_3\end{bmatrix}=\begin{bmatrix}-5\\ 0\\ 3\end{bmatrix}$$

再用如下 MATLAB 程序求解该方程组，即

```
≫A=[6 -2 -2; -2  9 -6; -4 -4  9];
≫F=[-5; 0; 3];
≫V=A\F
V=
-1.1471
-0.5294
-0.4118
```

可得 $V_3=-0.4118$V。■

如图 8.21 所示，如果在两个基节点之间的某条支路含有独立电压源或者受控电压源，那么，通过电源的电流就不能用节点电压来表示。此时，可以将这两个节点合成一个超节点，只要写出这个超节点的方程，用流出节点 2 的电流来表示流过电源的电流 I_A 就可以了。也就是，用节点电压写出 $I_B+I_C+I_D$ 来代替 I_A。由于有两个未知数，只有一个超节点方程，因此，需要对两个节点和电源应用 KVL 再写出一个方程，即

$$-V_1-V_\Delta+V_2=0$$

也就是

$$V_\Delta=V_2-V_1$$

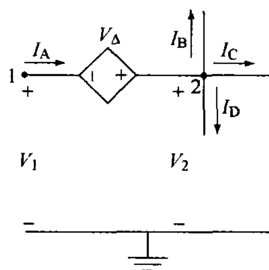
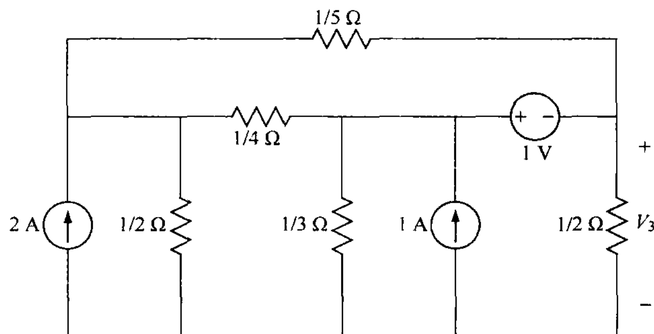
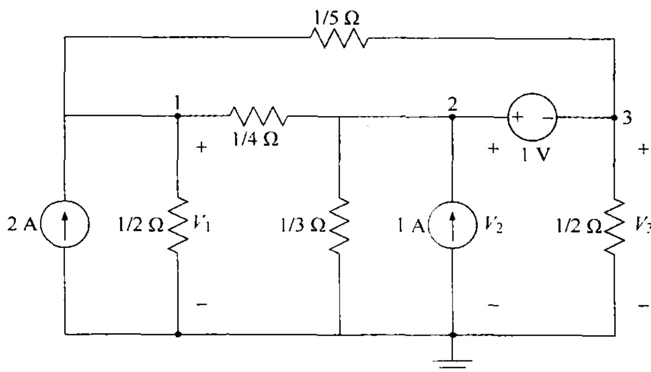


图 8.21　节点 1 和节点 2 之间存在受控电压源

例 8.7　请求出如下电路中的电压 V_3。



解：该电路有 3 个基节点，其中两个节点之间连接着一个独立电压源，形成超节点。在如下重画的电路图中，分别用“1”、“2”和“3”标出了这 3 个基节点，并且标出了电路底部的参考节点，以及 3 个节点电压 V_1、V_2和 V_3。



流出节点 1 的电流之和为

$$-2+2V_1+5(V_1-V_3)+4(V_1-V_2)=0$$

简化之后为

$$11V_1-4V_2-5V_3=2$$

节点 2 和节点 3 之间连接了一个独立电压源，因此形成一个超节点 2+3。流出这个超节点 2+3 的电流之和为

$$4(V_2-V_1)+3V_2-1+2V_3+5(V_3-V_1)=0$$

简化之后为

$$-9V_1+7V_2+7V_3=1$$

对 2、3 两个节点电压和电源应用 KVL 写出超节点的另一个方程

$$-V_2+1+V_3=0$$

也就是

$$-V_2+V_3=-1$$

将以上三个方程写成矩阵形式，即

$$\begin{bmatrix} 11 & -4 & -5 \\ -9 & 7 & 7 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$

用如下 MATLAB 程序求解该方程组

```
≫A=[11 −4 −5; −9  7  7; 0 −11];
≫F=[2; 1; −1];
≫V=A \ F
V=
0.4110
0.8356
```

-0.1644

可得 $V_3=-0.1644$。■

8.6 线性叠加原理

如果有两个或两个以上独立激励源同时作用于一个线性系统，那么，系统总的输出响应等于系统对于各个输入的响应之和。这种性质称为叠加原理。根据叠加原理，电路系统对于几个独立电源的总响应就等于每个独立电源响应之和。在求解各个独立电源的响应时，要去掉其他独立电源，也就是

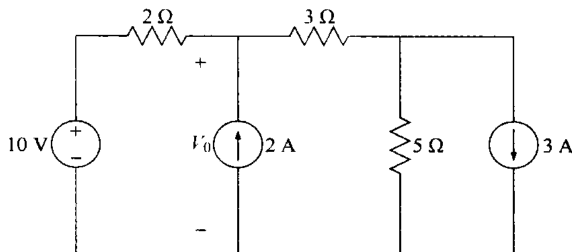
1）将其他电压源短路；

2）将其他电流源开路。

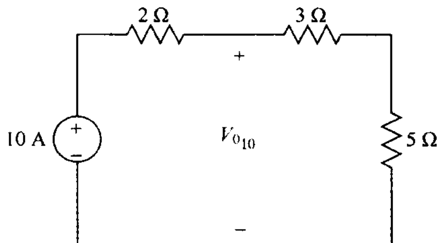
这样，在求第一个独立电源的响应时，去掉其他独立电源的作用。然后用同样的方法求下一个独立电源的响应。每个独立电源的响应都得到之后，将这些响应求和即可。但是，必须注意，叠加原理只适用于独立电源。在求各个独立电源的响应时，受控电源不能去掉，都必须保留，它们的值由相应的电流或者电压决定。显然，在求各个独立电源的响应时，电路不同，控制受控电源的那些电流和电压也不同，不能相互混用。

一般，与直接求取所有电源的总响应相比，利用叠加原理求解电路要简单得多。特别对于输入源为脉冲信号或者具有延时的信号，叠加原理很有用。在下面的章节中会介绍有关内容。

例 8.8　请应用叠加原理求如下电路中的电压 V_0。



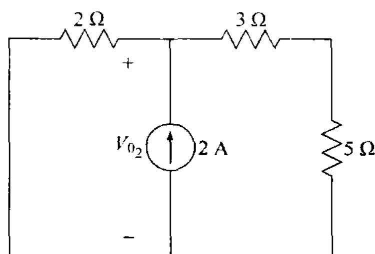
解：首先将两个电流源去掉，计算只有 10V 电压源作用下的响应，其电路如下：



根据分压器原理，可以方便地求得 10V 电压源所产生的 $V_{0_{10}}$，即

$$V_{0_{10}}=10\left(\frac{8}{2+8}\right)=8\text{V}$$

然后，求 2A 电流源作用下的响应，其电路如下：



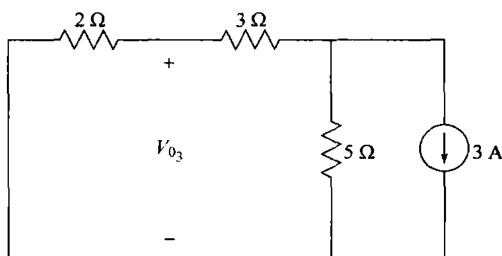
计算该电路的等效电阻，即

$$R_{EQ}=2\parallel(3+5)=\frac{2\times8}{2+8}=1.6\Omega$$

再应用欧姆定律，可得

$$V_{0_2}=2\times1.6=3.2\text{V}$$

最后，求 3A 电流源作用下的响应，其电路如下：



在该电路中，3A 电流源被两条支路（2+3Ω 和 5Ω）均分流成两个 1.5A 电流，于是

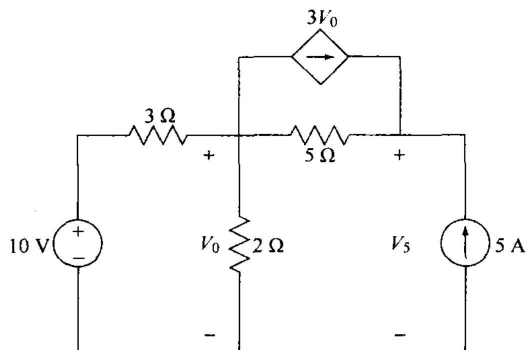
$$V_{0_3}=-1.5\times2=-3\text{V}$$

将以上各个响应求和，就得到总的 V_0 值：

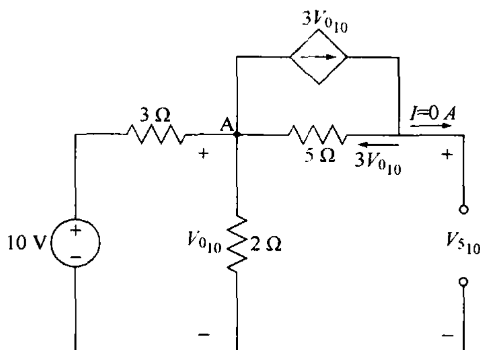
$$V_0=V_{0_{10}}+V_{0_2}+V_{0_3}=8+3.2-3=8.2\text{V}$$

■

例 8.9 请应用叠加原理求如下电路中 5A 电流源两端的电压 V_5。



解：首先，如下图所示，去掉 5A 电流源，计算 10V 电压源的响应 $V_{o_{10}}$。注意，受控电流源不能去掉，要保留。



该电路中去掉电流源的开路支路上没有电流，流过 5Ω 电阻的电流为 $3V_o$，因此，在节点 A 处应用 KCL，可得

$$\frac{V_{o_{10}}-10}{3}+\frac{V_{o_{10}}}{2}+3V_{o_{10}}-3V_{o_{10}}=0$$

于是

$$V_{o_{10}}=4\text{V}$$

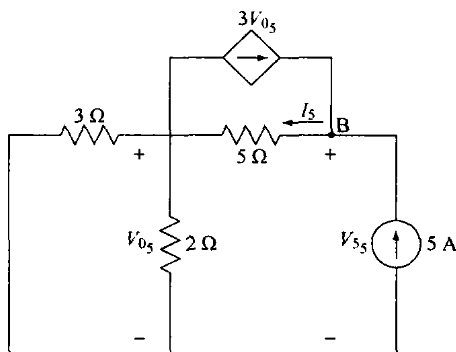
再应用 KVL，可得

$$-V_{o_{10}}-5\times 3V_{o_{10}}+V_{5_{10}}=0$$

于是

$$V_{5_{10}}=64\text{V}$$

然后，如下图所示，去掉 10V 电压源，计算 5A 电流源作用下的响应 V_{o_5}。



两个并联电阻（3Ω ∥ 2Ω）组合之后的等效电阻是 1.2Ω，由欧姆定律计算得到 $V_{o_5}=5\times 1.2=6\text{V}$。再在节点 B 处应用 KCL 计算 I_5，即

$$-3V_{o_5}+I_5-5=0$$

将 $V_{o_5}=6\text{V}$ 代入，可得 $I_5=3\times 6+5=23\text{A}$。最后，应用 KVL，有

$$-V_{o_5}-5I_5+V_{5_5}=0$$

于是

$$V_{5_3}=V_{0_5}+5I_5=6+5\times23=121\text{V}$$

再将各个响应求和，就得到总的响应

$$V_5=V_{5_{10}}+V_{5_5}=64+121=185\text{V}$$

■

8.7 戴维南定理

任何具有两个外接端点 A 和 B 并由电阻、受控电源和独立电源组成的电路都可以用一个电阻和一个独立电源来代替。如图 8.22 所示，由原电路简化得到的一个电压源和一个电阻串联的电路称为戴维南等效电路。连接在外接端 A 和 B 上的外接电路元件称为负载（图中没有显示）。戴维南等效电路提供给任何外接负载的电压和电流都与原电路相同。负载一般不包含在简化的等效电路之内，因为通常需要计算负载上的最大消耗功率等参数。虽然这里涉及的电路元件主要是电源和电阻，不过，戴维南定理适用于任何由线性元件组成的并具有两个端点的电路。

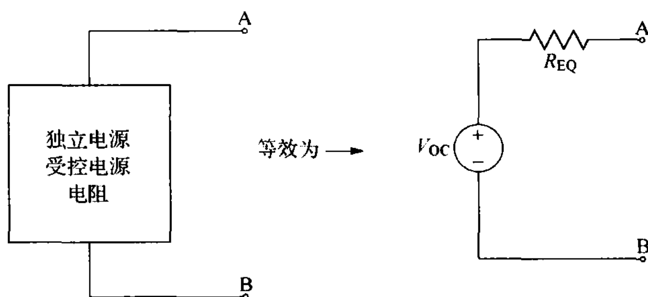


图 8.22 包含独立电源和受控电源的一般电路都可以等效为一个电压源 V_{OC} 与一个电阻 R_{EQ} 的串联

戴维南定理指出：由一个理想电压源 V_{OC} 和一个等效电阻 R_{EQ} 串联组成的等效电路可以代替任何由独立或受控电压源和电流源与电阻组成的电路。如图 8.23 所示，V_{OC} 等于 A、B 两端之间的开路电压，可以用节点电压等方法计算求得。电阻 R_{EQ} 是去掉所有电源之后从 A、B 两端看进去的等效电阻。记住，去掉电压源是用短路替代，去掉电流源则是用开路替代。

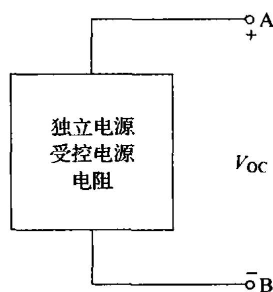
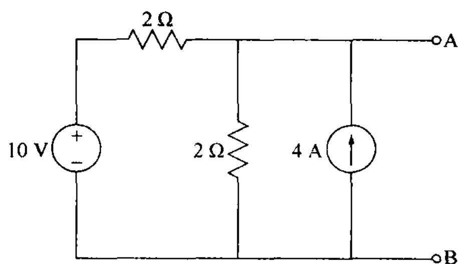


图 8.23 用节点电压法或回路电流法等方法计算得到 A、B 两端之间的开路电压 V_{OC}

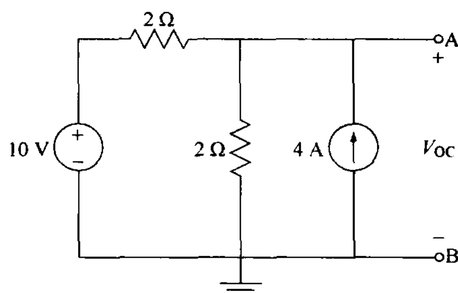
例 8.10 试求出以下电路相对于 A、B 两端的戴维南等效电路。



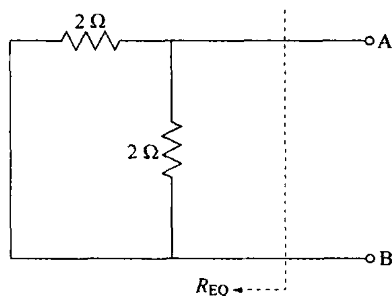
解： 戴维南等效电路分两步计算：求 V_{OC}；求 R_{EQ}。从以下电路可以方便地应用节点电压法计算得到开路电压 V_{OC}。即离开节点的电流之和为

$$\frac{V_{OC}-10}{2}+\frac{V_{OC}}{2}-4=0$$

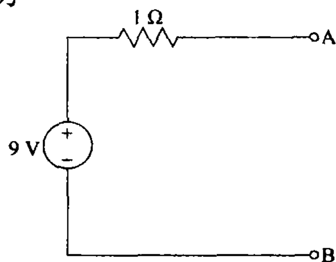
因此，$V_{OC}=9\text{V}$。



然后，如下图所示，去掉所有电源，也就是将电压源短路，将电流源开路。计算从 A、B 两端看进去的电阻 R_{EQ}。显然，R_{EQ}等于 1（即 2Ω ‖ 2Ω）。



因此，戴维南等效电路为



必须注意，用于计算 V_{OC} 的电路不能用于计算 R_{EQ}，两个计算电路是不同的，不能混淆。■

如图 8.24 所示，如果将 A、B 两端短路，短路电流为 I_{SC}，则以下关系成立：

$$R_{EQ}=\frac{V_{OC}}{I_{SC}} \tag{8.18}$$

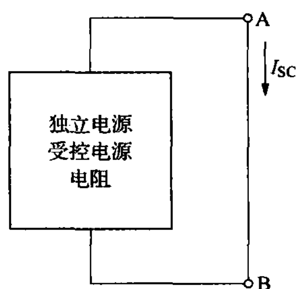


图 8.24 短路电路的等效电阻

8.8 电感

本章前几节介绍了可以用代数方程描述的包含电源和电阻的电路，这些电路中的任何电源变化都只产生即时响应。本节将介绍的电感也是一种无源元件，但其电压—电流关系需要用微分方程来描述。因此，含有电感的电路方程要用到导数和积分，电路中电源的任何变化（如阶跃输入）所产生的响应不是即时的，而是由两部分组成，一是按照指数变化的自由响应（natural response），二是与电源具有相同形式的强迫响应（forced response）。

电感是无源元件，它由绕在绝缘体或铁磁体心子上的导线线圈构成，可以将电能存储在磁场中。线圈中有电流流过时就产生了磁场。电路中的电感用符号⌒⌒⌒表示。电感的单位是亨利（H），1H=1V·s/A。电感的电压—电流关系是

$$V=L\frac{\mathrm{d}i}{\mathrm{d}t} \tag{8.19}$$

如图 8.25 所示，电感上电压降的表示方式通常与电阻相同。

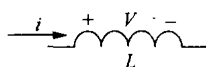
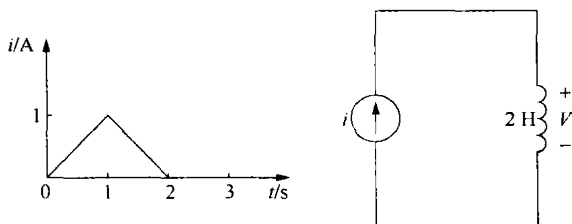


图 8.25 电感

由式（8.19）可知，实际的电感电流不可能即时变化，因为电流的即时变化会使其导数变成无穷大，从而产生无穷大的电压。但是，在数学上，使用 δ 函数（Dirac delta function）的电压就可以在电感上产生阶跃电流。按照惯例，如果电路中只存在直流电流或直流电压，那么，电感就可以看作短路，因为此时电感两端的电压差为 0。

例 8.11 请计算以下电路中的电压 v。



解：求解本题时，最好根据输入电流的变化将时间分段。显然，当 $t<0$ 和 $t>2$ 时，电流为 0，因此，$v=0$。下面利用式（8.19）求其他两个时间段上的电压值。

当 $0<t<1$ 时，输入电流为 $i=t$，则

$$v=L\frac{di}{dt}=2\frac{d(t)}{dt}=2V$$

当 $1\leqslant t\leqslant 2$ 时，输入电流为 $i=-(t-2)$，则

$$v=L\frac{di}{dt}=2\frac{d(-(t-2))}{dt}=-2V$$

式（8.19）是由给定电流求电感两端电压，如果反过来，给定电感两端的电压，需要计算电流，那么，将该方程两边同乘以 dt，得到

$$v(t)dt=Ldi$$

再两边求积分，即

$$\int_{t_0}^{t} v(\lambda)d\lambda = L\int_{i(t_0)}^{i(t)} d\alpha$$

就可以得到

$$i(t) = \frac{1}{L}\int_{t_0}^{t} v(\lambda)d\lambda + i(t_0) \tag{8.20}$$

通常设 $t_0=0$，于是，上式可以简化为

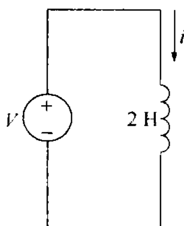
$$i(t) = \frac{1}{L}\int_{0}^{t} v(\lambda)d\lambda + i(0) \tag{8.21}$$

如果设 $t_0=-\infty$，且初始电流为 0，则式（8.20）可以简化为

$$i(t) = \frac{1}{L}\int_{-\infty}^{t} v(\lambda)d\lambda \tag{8.22}$$

式（8.20）中的初始电流 $i(t_0)$ 通常定义为与 i 的方向相同，也就是 $i(t_0)$ 为正值。如果 $i(t_0)$ 的方向与 i 的方向相反（写节点方程时会遇到这种情况），那么，$i(t_0)$ 就是负值。■

例 8.12　假设以下电路中 $i(0)=2A$，并且 $v(t)=4e^{-3t}u(t)$，请计算 $t\geqslant 0$ 时的 i。



解：由式（8.20）可得

$$i(t) = \frac{1}{L}\int_{t_0}^{t} v d\lambda + i(t_0) = \frac{1}{2}\int_{0}^{t} 4e^{-3\lambda}d\lambda + i(0) = \frac{1}{2}\int_{0}^{t} 4e^{-3\lambda}d\lambda + 2$$

$$=2\left.\frac{e^{-3\lambda}}{-3}\right|_{\lambda=0}^{t}+2$$

$$=\frac{2}{3}(4-e^{-3t})u(t)V$$

■

8.9 电容

电容是一种利用电场存储电能的装置，将电压加在电容上就可以使它极化，产生电荷分离，从而存储能量。简单的电容由两块导电材料制成的平行极板组成，极板之间填充了电介质。电介质可以是空气、云母、聚四氟乙烯等材料，这些材料含有很多电偶极子，在电场作用下会被极化。电介质极化引起的电荷分离与外加电压成正比，即

$$q(t)=Cv(t) \tag{8.23}$$

式中 C——电容，单位是 F，1F=1C/V。

本书用符号⊣⊢表示电容。一般电容的大小用微法（1μF=10^{-6}F）或者皮法（1pF=10^{-12}F）为单位来表示。图 8.26 就是一个包含电容的电路。

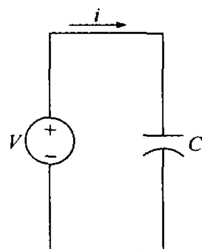


图 8.26 包含电容的电路

由式（8.23）可得如下更有用的电流—电压关系式

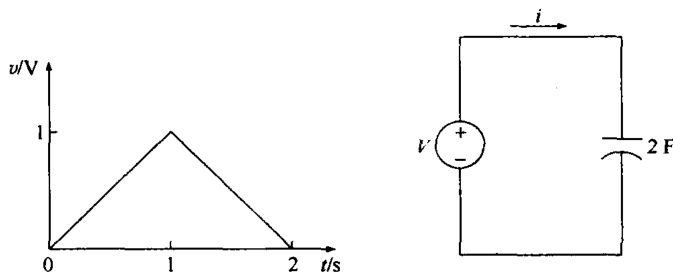
$$i=\frac{\mathrm{d}q}{\mathrm{d}t}=C\frac{\mathrm{d}v}{\mathrm{d}t} \tag{8.24}$$

电容的大小与两块极板之间的电介质的介电常数有关（例如空气的介电常数为 $\epsilon=8.854\times10^{-12}\frac{\mathrm{F}}{\mathrm{m}}$），并且与极板之间的距离 d 和极板的截面积 A 有关，即

$$C=\frac{\epsilon A}{d} \tag{8.25}$$

如上所述，实际的电容由有两块存储电荷的导电极板组成，两块极板之间由阻抗很大的绝缘材料薄层隔开，因此，电容极板之间实际上并没有电流流过。就像麦克斯韦（James Clerk Maxwell）在描述均匀电磁场时所提出的假说那样，在电容极板之间存在着一种内部电流转移，这种电流与流入和流出电容的电流相等，于是，KCL 仍然成立。显然，由式（8.24）可知电介质材料不传导直流电流。对于直流电，电容相当于开路。

例 8.13 请计算以下电路中的电流 i。



解：当 $t<0$ 和 $t>2$ 时，$v=0\text{V}$，因此，$i=0$。当 $0\leqslant t\leqslant 1$ 时，$v=t\text{V}$；当 $1<t\leqslant 2$ 时，$v=-(t-2)\text{V}$，利用式（8.24）可以计算这两个时间段上的电流。

当 $0\leqslant t\leqslant 1$ 时，有

$$i=C\frac{\mathrm{d}v}{\mathrm{d}t}=2\times\frac{\mathrm{d}}{\mathrm{d}t}(t)=2\text{A}$$

当 $1<t\leqslant 2$ 时，有

$$i=C\frac{\mathrm{d}v}{\mathrm{d}t}=2\times\frac{\mathrm{d}}{\mathrm{d}t}(-(t-2))=-2\text{A}$$

电容上的电压不能即时变化，要在电容上产生一个阶跃电压需要无限大的电流，这在实际电路中是不可能的。不过，在数学上可以利用 δ 函数来实现。

式（8.24）定义了给定电压下电容电流的计算公式，如果反过来，给定电容电流，要求计算电压，那么，将式（8.24）两边同乘以 $\mathrm{d}t$，可得

$$i(t)\mathrm{d}t=C\mathrm{d}v$$

此式两边求积分，可得

$$\int_{t_0}^{t} i(\lambda)\mathrm{d}\lambda = C\int_{v(t_0)}^{v(t)}\mathrm{d}v$$

也就是

$$v(t) = \frac{1}{C}\int_{t_0}^{t} i(\lambda)\mathrm{d}\lambda + v(t_0) \tag{8.26}$$

如果设 $t_0=0$，此式可以简化为

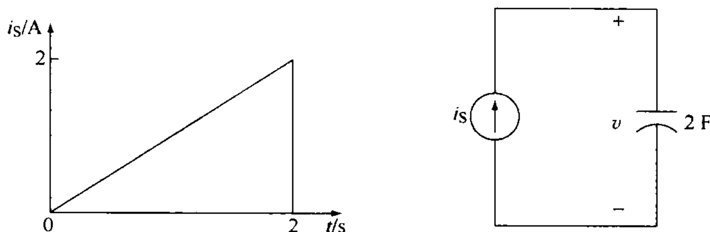
$$v(t) = \frac{1}{C}\int_{0}^{t} i(\lambda)\mathrm{d}\lambda + v(0) \tag{8.27}$$

如果设 $t_0=-\infty$，则可以简化为

$$v(t) = \frac{1}{C}\int_{-\infty}^{t} i(\lambda)\mathrm{d}\lambda \tag{8.28}$$

式（8.26）中的初始电压 $v(t_0)$ 通常定义为与 v 的极性相同，也就是 $v(t_0)$ 为正值。如果 $v(t_0)$ 的极性与 v 的极性相反，那么，$v(t_0)$ 就是负值。 ■

例 8.14　请计算以下电路中的电压 v。



解：本题的电流波形可以分成 3 个时间段：$t\leqslant 0$，$0<t\leqslant 2$，$t>2$。应用式（8.28）可以求出各个时间段的电压。

当 $t\leqslant 0$ 时

$$v(t)=\frac{1}{C}\int_{-\infty}^{t} i\mathrm{d}t=\frac{1}{2}\int_{-\infty}^{0} 0\mathrm{d}t=0\mathrm{V}$$

当 $0<t\leqslant 2$ 时

$$v(t)=\frac{1}{C}\int_{0}^{t} i\mathrm{d}t+v(0)$$

如果设 $v(0)=0$，则

$$v(t)=\frac{1}{2}\int_{0}^{t}\lambda\mathrm{d}\lambda=\frac{1}{2}\left(\frac{\lambda^{2}}{2}\right)\Big|_{0}^{t}=\frac{t^{2}}{4}\mathrm{V}$$

$t=2$ 时的电压值为（此值是下一步计算的初始值）

$$v(2)=\frac{t^{2}}{4}\Big|_{t=2}=1\mathrm{V}$$

当 $t>2$ 时

$$v(t)=\frac{1}{C}\int_{2}^{t} i\mathrm{d}t=v(2)=\frac{1}{2}\int_{2}^{t} 0\mathrm{d}t+v(2)=1\mathrm{V}$$ ■

8.10 电阻、电容和电感组成的电路的通用求解方法

有时，由电阻、电容和电感组成的电路不能用串联组合和并联组合的方式进行简化。如图 8.27 所示的电路就没有电阻、电容或电感的串联和并联，不能通过简化电路来求解。本节将通过节点电压法，利用电阻、电容和电感的电压—电流关系式，写出包含积分和微分的电路方程，并用标准微分方程求解法解这些方程，来获得各个未知的电压和电流。

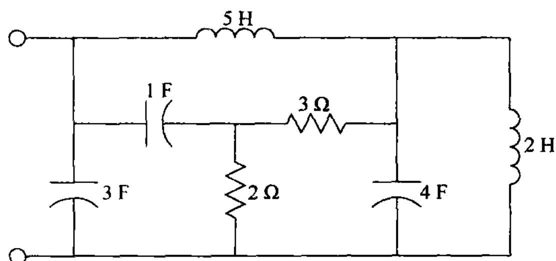
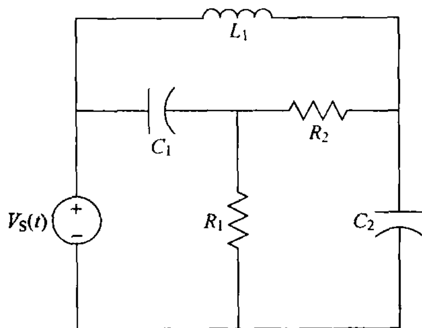


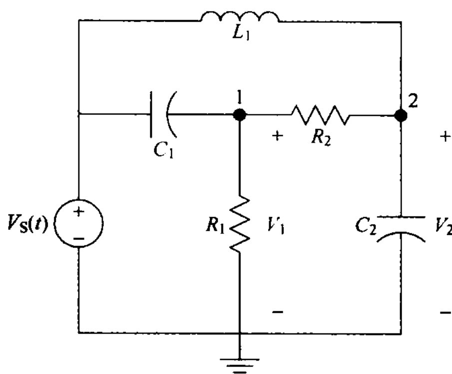
图 8.27 不能简化的电路

例 8.15 假设初始条件为 0，请写出以下电路在 $t\geqslant 0$ 时的节点方程。



解：在如下重画的电路图中，将电路底部的节点设为参考节点，另外有“1”和

“2”两个基节点。记住，与电压源相连的节点电压已知，不需要写节点方程。在写节点方程时，电容电流为 $i_c=C\Delta\dot{v}$，其中 $\Delta\dot{v}$ 为电容两端电压的导数；电感电流为 $i_L=\frac{1}{L}\int_0^t \Delta v \mathrm{d}\lambda+i_L(0)$，其中 Δv 为电感两端的电压。由于初始值为 0，因此，$i_L(0)=0$。



离开节点 1 的电流之和为

$$C_1(\dot{v}_1-\dot{v}_S)+\frac{v_1}{R_1}+\frac{v_1-v_2}{R_2}=0$$

可以简化为

$$C_1\dot{v}_1+\left(\frac{1}{R_1}+\frac{1}{R_2}\right)v_1-\frac{1}{R_2}v_2=C_1\dot{v}_S$$

离开节点 2 的电流之和为

$$\frac{v_2-v_1}{R_2}+C_2\dot{v}_2+\frac{1}{L_1}\int_0^t(v_2-v_S)\mathrm{d}\lambda=0$$

通过微分可以将上式节点方程中的积分去掉，变成

$$\frac{1}{R_2}\dot{v}_2-\frac{1}{R_2}\dot{v}_1+C_2\ddot{v}_2+\frac{1}{L_1}v_2-\frac{1}{L_1}v_S=0$$

重排此式，可得

$$\ddot{v}_2+\frac{1}{C_2R_2}\dot{v}_2+\frac{1}{C_2L_1}v_2-\frac{1}{C_2R_2}\dot{v}_1=\frac{1}{C_2L_1}v_S$$ ■

使用节点电压法时，每个基节点产生一个方程。我们可以利用其他节点方程，通过微分以及各种代换，将目标节点的方程转化为只包含输入函数和一个未知节点电压的单变量微分方程。还有一种建立单变量微分方程的方法是利用 D 算子，详见后面的例 8.17。

例如，对于例 8.15 的电路，要求写出节点电压 v_1 及其导数组成的单变量微分方程。为了方便，假设各个电路元件的值为：$R_1=R_2=1\Omega$，$C_1=C_2=1\text{F}$，$L_1=1\text{H}$，则有如下两个方程：

$$\dot{v}_1+2v_1-v_2=\dot{v}_S$$

和

$$\ddot{v}_2+\dot{v}_2+v_2-\dot{v}_1=v_S$$

从第一个方程求得 v_2，并计算$\dot{v}_2$ 和$\ddot{v}_2$，即

$$v_2=\dot{v}_1+2v_1-\dot{v}_S$$

$$\dot{v}_2=\ddot{v}_1+2\dot{v}_1-\ddot{v}_S$$

$$\ddot{v}_2=\dddot{v}_1+2\ddot{v}_1-\dddot{v}_S$$

然后代入第二个方程，则有

$$\dddot{v}_1+2\ddot{v}_1-\dddot{v}_S+\ddot{v}_1+2\dot{v}_1-\ddot{v}_S+\dot{v}_1+2v_1-\dot{v}_S-\dot{v}_1=v_S$$

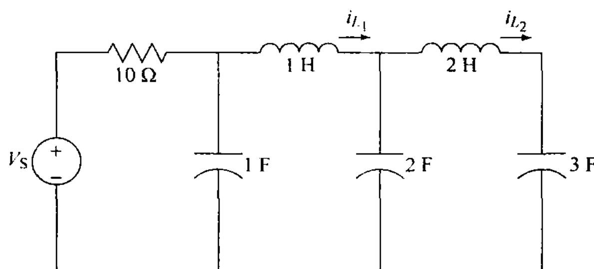
简化后，得到

$$\dddot{v}_1+3\ddot{v}_1+2\dot{v}_1+2v_1=\dddot{v}_S+\ddot{v}_S+\dot{v}_S-v_S$$

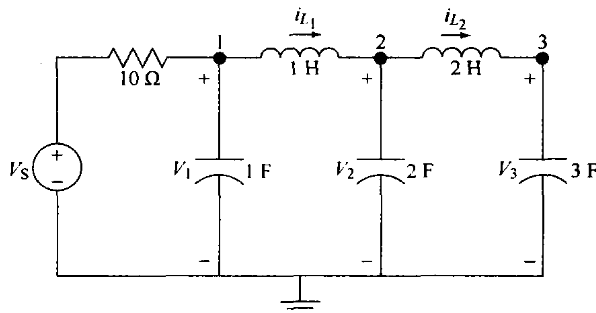
一般，由单个输出变量和各个输入组成的微分方程的阶数等于电路中储能元件（即电容和电感）的个数。但是，如果电容电压和电感电流不独立，也就是电容电压或者电感电流与输入之间存在直接的代数关系，那么，电路微分方程的阶数就小于电路中电容和电感的个数。当电容直接与电压源相连或者电感直接与电流源相连时就会出现这种情况。

上述例题有一个条件，就是假设电路的初始值为 0。如果电路有非 0 初始值，那么，用同样的方法也可以求解。这时只要在节点电压方程中加上电感的初始电流即可，如下例所示。

例 8.16　假设初始条件为 $i_{L_1}(0)=8\text{A}$ 和 $i_{L_2}(0)=-4\text{A}$，请写出以下电路 $t\geqslant0$ 时的节点方程。



解：如下重画电路图中，将电路底部的节点设为参考节点，另外有“1”、“2”和“3” 3 个基节点。



离开节点 1 的电流之和为

$$\frac{(v_1 - v_S)}{10} + \dot{v}_1 + \int_0^t (v_1 - v_2)\mathrm{d}\lambda + 8 = 0$$

其中 $i_{L_1}(0)=8\mathrm{A}$。

离开节点 2 的电流之和为

$$\int_0^t (v_2 - v_1)\mathrm{d}\lambda - 8 + 2\,\dot{v}_2 + \frac{1}{2}\int_0^t (v_2 - v_3)\mathrm{d}\lambda - 4 = 0$$

其中 $i_{L_2}(0)=-4\mathrm{A}$。注意，由于该初始电感电流的方向是从右到左，与电路图中定义的方向相反，因此，电流为负值。

离开节点 3 的电流之和为

$$\frac{1}{2}\int_0^t (v_3 - v_2)\mathrm{d}\lambda + 4 + 3\,\dot{v}_3 = 0$$

由于将方程微分去掉积分时，会消去方程中的初始电感电流，因此，本例的节点方程没有简化。如果需要写出只包含一个节点电压以及输入的单变量微分方程，由于电路中有 5 个储能元件，因此会产生 5 阶微分方程。要求解这个 5 阶微分方程，就需要 5 个初始条件，也就是在 $t=0$ 时刻所选定节点电压的初始值及其 1～4 阶导数值。■

电路的突变和初始条件

电路中有阶跃输入或者有开关动作时，就会产生电压和电流的突变。我们已经知道，求解 n 阶微分方程必须有 n 个初始条件，这些初始条件一般就是阶跃输入时刻或者开关动作时刻输出变量及其 $n-1$ 个导数的值。如果整个时间段上电路的各个输入都已知，那么，基于储能元件的能量不能瞬时突变的条件，就可以直接求得初始值，而不需要在题目中事先给定。几乎所有电路问题都可以假设输入发生在 $t=0$ 时刻，因此，下面主要讨论 $t=0$ 时刻的输入。不过，这些方法也很容易推广到任意时刻的输入。

储能元件的能量不能瞬时突变，所以，电感的电流或者电容的电压不会突变，在 $t=0^-$ 和 $t=0^+$ 时这些变量保持同样的值。例如，例 8.16 中，给定电感和电容的初始条件时，就意味着 $i_{L_1}(0^-)=i_{L_1}(0^+)$、$i_{L_2}(0^-)=i_{L_2}(0^+)$；以及 $v_1(0^-)=v_1(0^+)$、$v_2(0^-)=v_2(0^+)$、$v_3(0^-)=v_3(0^+)$。除了与电感电流和电容电压有关的变量以外，其他变量在阶跃输入或者开关动作时可以产生突变；但是，这些变量的变化必须遵循 KVL 和 KCL。

虽然不是一眼就可以看得出来，但是，在 $t=0^-$ 和 $t=0^+$ 时，电感电流和电容电压的导数可以发生突变，因为既然 $v_L(0^+)$ 和 $i_C(0^+)$ 可以突变，就有

$$\frac{\mathrm{d}i_L(0^+)}{\mathrm{d}t}=\frac{v_L(0^+)}{L},\quad \frac{\mathrm{d}v_C(0^+)}{\mathrm{d}t}=\frac{i_C(0^+)}{C}$$

记住，要在这些表达式求导之后，再求导数在 $t=0$ 时刻的值，也就是

$$\frac{\mathrm{d}i_L(0^+)}{\mathrm{d}t}=\left.\frac{\mathrm{d}i_L(t)}{\mathrm{d}t}\right|_{t=0^+},\quad \frac{\mathrm{d}v_C(0^+)}{\mathrm{d}t}=\left.\frac{\mathrm{d}v_C(t)}{\mathrm{d}t}\right|_{t=0^+}$$

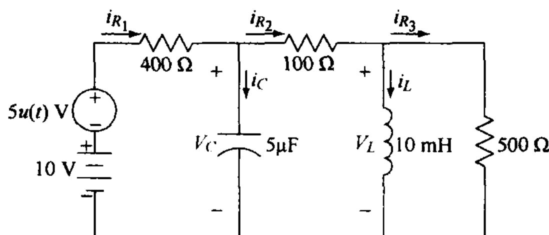
在计算与电感电流和电容电压无关的那些变量的导数时，可能需要用到阶跃输入的导数值，这里假设 $t=0^+$ 时阶跃输入的导数为 0。

与电感电流和电容电压无关的那些变量在突变时刻的初始条件完全由与电感电流和电容电压相关的变量以及所施加的输入电源决定，利用 KCL 和 KVL，或者利用节点电压法，这些初始值的计算可以分为如下两步：

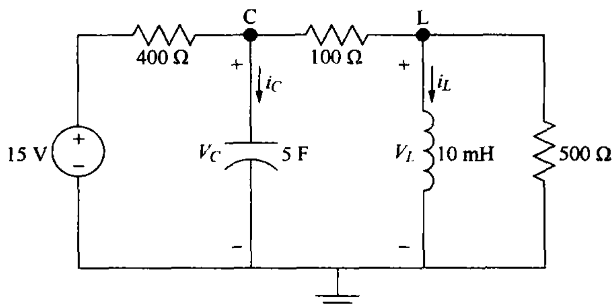
1）分析 $t=0^-$ 时的电路。记住，电路处于稳态时，电感相当于短路，电容相当于开路。因此，在稳态的 $t=0^-$ 时刻，将电路中所有电感短路，将所有电容开路。然后，计算流过电感的电流和电容两端的电压。

2）分析 $t=0^+$ 时的电路。从 $t=0^-$ 到 $t=0^+$ 时，由于电感电流不变，因此，可以用电流值为 $t=0^-$ 时值的电流源来代替电感。同理，从 $t=0^-$ 到 $t=0^+$ 时，由于电容电压不变，可以用电压值为 $t=0^-$ 时值的电压源来代替电容。于是，利用这样的电路，就可以计算出所有求解微分方程所需要的初始条件。

例 8.17 请利用节点电压法计算以下电路在 $t>0$ 时的电压 v_C。



解：将 $t>0$ 时的电路重画如下



离开节点 C 的电流之和为

$$\frac{v_C-15}{400}+5\times10^{-6}\dot{v}_C+\frac{v_C-v_L}{100}=0$$

可以简化为

$$\dot{v}_C+2500v_C-2000v_L=7500$$

离开节点 L 的电流之和为

$$\frac{v_L-v_C}{100}+\frac{1}{10\times10^{-3}}\int_0^t v_L\mathrm{d}\lambda+i_L(0^+)+\frac{v_L}{500}=0$$

将其乘以 500 并微分，可得

$$6\dot{v}_L+50\times10^3v_L-5\dot{v}_C=0$$

利用 D 算子，将两个微分方程写成

$$Dv_C+2500v_C-2000v_L=7500,\text{即}(D+2500)v_C-2000v_L=7500$$

$$6Dv_L+50\times10^3v_L-5Dv_C=0,\text{即}(6D+50\times10^3)v_L-5Dv_C=0$$

于是，由第一个方程可以解得 v_L

$$v_L=(0.5\times10^{-3}D+1.25)v_C-3.75$$

将此 v_L 代入第二个方程，可得

$$(6D+50\times10^3)v_L-5Dv_C=(6D+50\times10^3)((0.5\times10^{-3}D+1.25)v_C-3.750)-5Dv_C=0$$

简化之后，可得

$$D^2v_C+10.417\times10^3Dv_C+20.83\times10^6v_C=62.5\times10^6$$

再转化回到时域，即

$$\ddot{v}_C+10.417\times10^3\dot{v}_C+20.83\times10^6v_C=62.5\times10^6$$

该微分方程的特征方程为

$$s^2+10.417\times10^3s+20.833\times10^6=0$$

其根是-7.718×10^3和-2.7×10^3，于是，微分方程的自由解是

$$v_{C_n}(t)=K_1\mathrm{e}^{-7.718\times10^3t}+K_2\mathrm{e}^{-2.7\times10^3t}\mathrm{V}$$

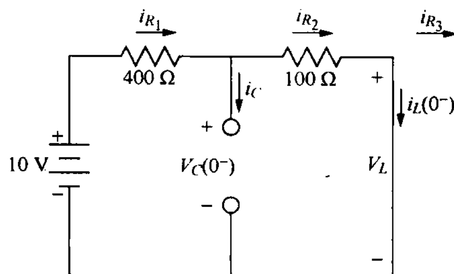
下面求微分方程的强迫响应。假设 $v_{C_f}(t)=K_3$，将其代入微分方程，有

$$20.833\times10^6K_3=62.5\times10^6$$

即 $K_3=3$。于是，方程的解就是

$$v_C(t)=v_{C_n}(t)+v_{C_f}(t)=K_1\mathrm{e}^{-7.718\times10^3t}+K_2\mathrm{e}^{-2.7\times10^3t}+3\mathrm{V}$$

求 K_1和 K_2时需要初始条件 v_C（0^+）和$\dot{v}_C(0^+)$，它们的值求解如下。当 $t=0^-$时，将电路中的电容开路并将电感短路，得到如下电路：



注意，由于电感短路，此时 $v_L(0^-)=0$。并且 500Ω 的电阻也被短路，图中已省略，因此 $i_{R_3}(0^-)=0$。利用分压器原理，有

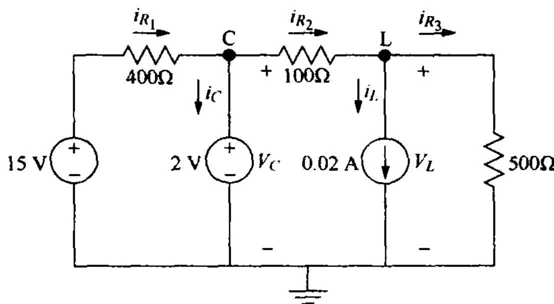
$$v_C（0^-）=10\times\frac{100}{400+100}=2\mathrm{V}$$

由欧姆定律可得

$$i_L（0^-）=\frac{10}{100+400}=0.02\mathrm{A}$$

由此可知 $i_{R_1}(0^-)=i_{R_2}(0^-)=i_L(0^-)=0.02\text{A}$。从 $t=0^-$ 到 $t=0^+$ 时，电容电压和电感电流不能突变，因此，$v_C(0^+)=v_C(0^-)=2\text{V}$，$i_L(0^-)=i_L(0^+)=0.02\text{A}$。

当 $t=0^+$ 时，如下图所示，将原电路中的电感用电流值等于 $t=0^-$ 时的电感电流值的电流源代替，将电容用电压值等于 $t=0^-$ 时的电容电压值的电压源代替。图中标出了参考节点和基本节点 C 和 L。注意，此时的输入还是 $10+5u(t)=15\text{V}$。



离开节点 L 的电流之和为

$$\frac{v_L-2}{100}+0.02+\frac{v_L}{500}=0$$

可得 $v_L(0^+)=0\text{V}$。现在，已知

$$i_{R_3}(0^+)=\frac{v_L(0^+)}{500}=0\text{A},\quad i_{R_2}(0^+)=0.02+i_{R_3}(0^+)=0.02\text{A},$$

$$i_{R_1}(0^+)=\frac{15-2}{400}=0.0325\text{A}。$$

要计算 $i_C(0^+)$，对节点 C 应用 KCL，可得

$$-i_{R_1}(0^+)+i_C(0^+)+i_{R_2}(0^+)=0$$

也就是

$$i_C(0^+)=i_{R_1}(0^+)-i_{R_2(0^+)}=0.0325-0.02=0.125\text{A}$$

下面计算 $\dot{v}_C(0^+)$。注意 $i_C(0^+)=C\dot{v}_C(0^+)$，也就是

$$\dot{v}_C(0^+)=\frac{i_C(0^+)}{C}=\frac{0.0125}{5\times10^{-6}}=2.5\times10^3\ \frac{\text{V}}{\text{s}}$$

利用初始条件，求常数 K_1 和 K_2 如下：

$$v_C(0)=2=K_1+K_2+3$$

然后

$$\dot{v}_C(t)=-7.718\times10^3K_1\text{e}^{-7.718\times10^3t}-2.7\times10^3K_2\text{e}^{-2.7\times10^3t}$$

并且当 $t=0$ 时有

$$\dot{v}_C(0)=2.5\times10^3=-7.718\times10^3K_1-2.7\times10^3K_2$$

由此可得 $K_1=0.04$，$K_2=-1.04$。将这些值代入，可得 $t\geqslant0$ 时方程的解为

$$v_C(t)=0.04\text{e}^{-7.718\times10^3t}-1.04\text{e}^{-2.7\times10^3t}+3\text{V}$$ ■

8.11 运算放大器

前面 8.3 节曾经介绍过受电路中其他电压或者电流控制的受控电压源和受控电流源，它们都是双端器件。本节将介绍的运算放大器，简称为运放，是一种多端电子器件，它由很多晶体管、电阻和电容组成。其详细工作原理涉及二极管和晶体管的基本知识，本书不作介绍。不过，运算放大器在电路中的工作方式其实就像已经讲过的受控电压源。在生物医学仪器中，从放大器到滤波器，运放电路都是基础。用于生物医学信号检测的放大器具有非常高的输入阻抗，这样从被测系统中分流的测量电流才会很少，对被测系统的影响也就小。生物体的信号一般都很微弱，例如，心电图（ECG）的幅值为 mV 级，而脑电图（EEG）的幅值只有 μV 级。另外，仪器中一般都要用模拟滤波器去除信号中的噪声，这种滤波器通常用频域法设计。8.13 节将介绍模拟滤波器，有关滤波器的基础理论则在第 10 章讲述。

顾名思义，运算放大器就是一个放大器。但是，当它与其他电路元件组合在一起时，却能构成积分器、微分器、加法器和减法器等电路。图 8.28 所示是一种具有 8 个端口引线的双列封装（Dual-In-line Package，DIP）的运算放大器早期产品。



图 8.28 一种 8 端运算放大器

注：其中，NC 端是空的，没有连接。两个偏置调零用于器件的校正（一般不用），V+和V−是给电路供电的电源端。记住，V+和 V−都是相对于电路中共用接地点的电压。新型运算放大器有 10 个或者更多的连接端。

与前面所介绍的电路元件不同，每个运放有两个输入端和一个输出端，其符号表示如图 8.29所示，两个输入端分别为同相输入端（+）和反相输入端（−），两个电源端分别标为 V+和 V−。不过，运放符号的电源端常被省略，因为，除非运放运行于饱和状态，否则其电源一般不会影响运放电路的性质。

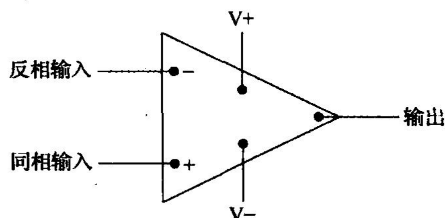


图 8.29 运算放大器的表示符号

图 8.30 所示是表示运放输入端和输出端

内部特性的模型，其输入与输出之间的关系为

$$V_o = A(V_p - V_n) \tag{8.29}$$

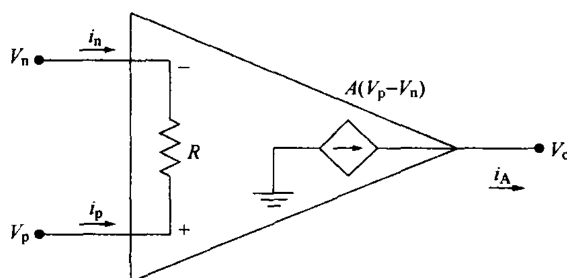


图 8.30　运放内部模型

注：两个输入端之间的内阻 R 非常大，超过 1MΩ。放大器的增益 A 也很大，超过 10^4。为了简化，图中省略了电源端。

由于运放输入阻抗很大，为了便于分析，可以将输入端视为开路，这样，就有了图 8.31 所示的运放模型。

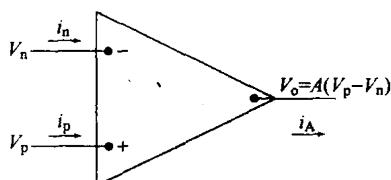
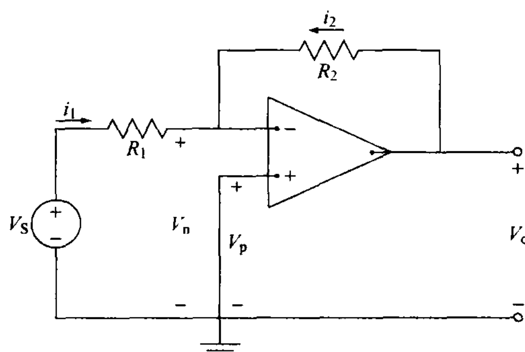


图 8.31　内阻 R 视为开路的理想运放模型

运放输入端的内阻 R 用开路代替之后，就有电流 $i_n = i_p = 0$A。但是，流出运放的电流 i_A 并不为 0。由于 i_A 未知，输出节点上一般不能用 KCL，因此，在运放电路求解中 KCL 总是用于输入端，详见以下例题。

例 8.18　请计算以下电路中的输出电压 V_0。



解： 由图 8.31 的运放模型可知，运放输入端没有电流流过，因此，在反相输入端应用 KCL，可得

$$-i_1-i_2=0$$

根据欧姆定律求得电流 i_1 和 i_2 的表达式并代入上式，即得

$$\frac{V_S-V_n}{R_1}+\frac{V_o-V_n}{R_2}=0$$

将此式两边同乘以 R_1R_2，并且合并同类项之后，可得

$$R_2V_S=(R_1+R_2)V_n-R_1V_o$$

而 $V_o=A(V_p-V_n)$，并且运放的同相输入端接地，即 $V_p=0$，因此

$$V_o=-AV_n$$

也就是

$$V_n=-\frac{V_o}{A}$$

将此 V_n 代入以上反相输入端的 KCL 方程，可得

$$R_SV_S=(R_1+R_2)\left(-\frac{V_o}{A}\right)-R_1V_o$$

$$=\left(\frac{R_1+R_2}{A}+R_1\right)V_o$$

即

$$V_o=-\frac{R_2V_S}{\left(R_1+\frac{R_1+R_2}{A}\right)}$$

当 A 趋于无穷大时，此式变成

$$V_o=-\frac{R_2}{R_1}V_S$$

有趣的是，由于电阻 R_2 的作用，即使 A 趋于无穷大，V_o 仍然为有限值。这是由于电阻 R_2 在输出端与反相输入端之间建立了一条负反馈通路，这种运放电路称为反相放大器，其总增益等于 $-\frac{R_2}{R_1}$。■

具有无穷大增益的运放称为理想运放。由于增益无穷大，因此，在运放的输出端和输入端之间必须有反馈通路，不能直接将输入电压连接在两个输入端之间。分析理想运放电路时，假设 $V_n=V_p$ 可以简化电路的分析计算。例如，对于以上例题 8.18，由于 $V_p=0$，$V_n=0$，在反相输入端应用 KCL，可得

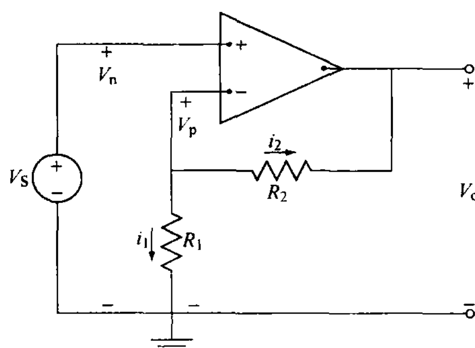
$$-\frac{V_S}{R_1}+\frac{-V_o}{R_2}=0$$

也就是

$$V_o=-\frac{R_2}{R_1}V_S$$

那么，为什么会有 $V_n=V_p$ 呢？记住，只要增益 A 足够大，并且存在反馈通路，这种假设就能够成立。

例 8.19　请计算以下电路的总增益。



解：假设运放是理想的，即 $V_n=V_p$。由于运放的同相端与输入电压相连，$V_n=V_p=V_S$，运放输入端无电流流入，因此，由 KCL 可得

$$i_1+i_2=0$$

也就是

$$\frac{V_S}{R_1}+\frac{V_S-V_o}{R_2}=0$$

即

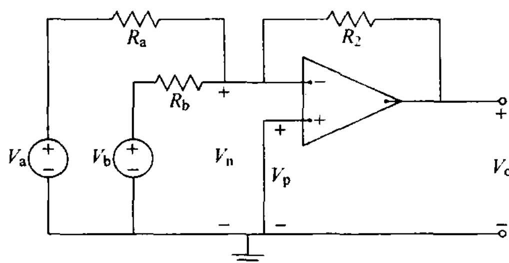
$$V_o=\left(\frac{R_1+R_2}{R_1}\right)V_S$$

因此，该电路的总增益为

$$\frac{V_o}{V_S}=\frac{R_1+R_2}{R_1}$$

这是一个用于放大输入信号的同相放大电路，心电图、脑电图、眼电图等临床检测仪器中都要用到这类放大器。下面举例说明加法运放电路。 ■

例 8.20 请计算以下电路的总增益。



解：还是从 $V_n=V_p$ 开始。注意，该运放的同相输入端与地相连，因此，$V_n=V_p=0\text{V}$。在反相输入端节点处应用 KCL，可得

$$-\frac{V_a}{R_a}-\frac{V_b}{R_b}-\frac{V_o}{R_2}=0$$

也就是

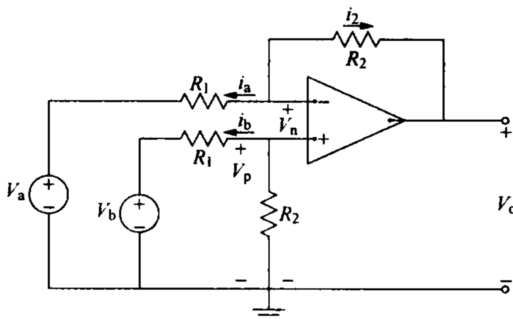
$$V_o=-\left(\frac{R_2}{R_a}V_a+\frac{R_2}{R_b}V_b\right)$$

可见，该电路是各个输入电压的加权求和。这种加法器的一般计算公式是

$$V_o=-\left(\frac{R_2}{R_a}V_a+\frac{R_2}{R_b}V_b+\cdots+\frac{R_2}{R_m}V_m\right)$$ ■

下面例题中的运放电路的输出与两个输入电压之差成正比，这种运放电路称为差动放大器。

例 8.21　请计算以下电路的总增益。



解：假设这是理想运放，没有电流流入运放的输入端，并且 $V_n=V_p$。在反相输入端节点处应用 KCL，可得

$$i_a=-i_2$$

也就是

$$\frac{V_n-V_a}{R_1}+\frac{V_n-V_o}{R_2}=0$$

即

$$(R_1+R_2)V_n-R_2V_a=R_1V_o$$

该方程有两个未知数，需要另加一个方程才能求解。在同相输入端应用分压器原理可得

$$V_p=\frac{R_2}{R_1+R_2}V_b=V_n$$

将其代入以上反相输入端的 KCL 方程，可得

$$R_2V_b-R_2V_a=R_1V_o$$

即

$$V_o=\frac{R_2}{R_1}(V_b-V_a)$$

可见，这种差动放大器的输出与两个输入信号之差成正比，它常用于心电图和脑电图等信号的双电极检测，这种检测方式有两个输入信号。理论上，差动放大器

只检测所需信号，可以除去环境噪声。因为，由照明、60Hz 供电线路[⊖]、接地不良以及电源漏电流等因素引起的噪声一般同时出现在两个输入端上，是共模信号，差动放大器可以抑制共模信号；而所需检测的信号则是两个输入端信号之差，差动放大器对其具有放大作用，详见下面的分析。 ■

差动放大器的输出响应可以分成差模（differential-mode）和共模（common-mode）两种成分，它们的定义分别是

$$V_{dm}=V_b-V_a$$

和

$$V_{cm}=\frac{(V_a+V_b)}{2}$$

可见，共模信号 V_{cm} 是两个输入端信号的平均。如果用 V_{dm} 和 V_{cm} 来表示两个输入端的电压 V_a 和 V_b，则

$$V_a=V_{cm}-\frac{V_{dm}}{2}$$

且

$$V_b=V_{cm}+\frac{V_{dm}}{2}$$

将它们代入以上例 8.21 的输出响应方程中，则

$$V_o=\left(\frac{R_1R_2-R_1R_2}{R_1(R_1+R_2)}\right)V_{cm}+\left(\frac{R_2(R_1+R_2)+R_2(R_1+R_2)}{2R_1(R_1'+R_2)}\right)V_{dm}=A_{cm}V_{cm}+A_{dm}V_{dm}$$

由于理想运放的共模增益 A_{cm} 为 0，此式中的共模信号项 $A_{cm}V_{cm}$ 也为 0，因此，理想运放只放大差模信号。但是，实际使用的运放是非理想的，并且电路中各个电阻的实际值也存在误差，因此，其共模增益 A_{cm} 并不等于 0。所以，设计差动放大器时，要尽可能减小共模增益 A_{cm}，增大差模增益 A_{dm}。

上述共模信号的抑制称为共模抑制（common-mode rejection），共模抑制比（Common-Mode Rejection Ratio，CMRR）是差动放大器的一个技术指标，其计算公式为

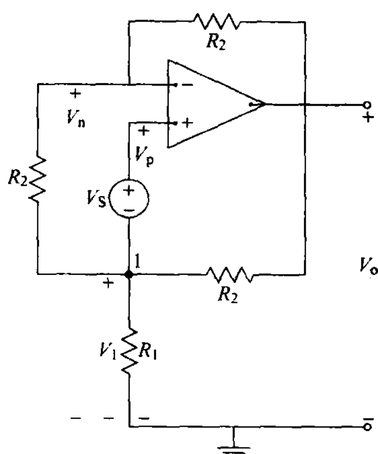
$$\text{CMRR}=20\lg\left|\frac{A_{dm}}{A_{cm}}\right|$$

CMRR 的值越大越好。用于脑电、心电和肌电等信号检测的差动放大器的共模抑制比一般为 100～120dB。

总之，运放电路求解的一般方法是：先假设运放是理想的，即 $V_n=V_p$；然后，在两个输入端应用 KCL 或者 KVL。如果电路比较复杂，那么，还要使用其他电路分析方法，下面举例说明。

例 8.22 请计算以下电路的输出电压 V_o。

⊖ 美国供电系统用的是 60Hz 交流电，中国和欧洲国家用的是 50Hz 交流电。——译者注



解：利用 $V_n=V_p$，在反相输入端应用 KCL，可得

$$\frac{V_n-V_1}{R_2}+\frac{V_n-V_o}{R_2}=0$$

即

$$2V_n-V_1-V_o=0$$

然后，在“地—节点 1—同相输入端—地”这个回路中应用 KVL，可得

$$-V_1-V_S+V_p=0$$

由于 $V_n=V_p$，因此，$V_n-V_1=V_S$。

现在再对节点 1 应用 KCL 定律，注意同相输入端没有电流流入，因此

$$\frac{V_1}{R_1}+\frac{V_1-V_o}{R_2}+\frac{V_1-V_n}{R_2}=0$$

合并式中的同类项之后，此式变为

$$-R_1V_n+(2R_1+R_2)V_1-R_1V_o=0$$

以上 3 个方程共包含 3 个未知数，可以求解。首先，由反相输入端的 KCL 方程与 KVL 方程求得 V_1：

$$V_1=V_o-2V_S$$

然后，将此 V_1 代入反相输入端的 KCL 方程，得到 V_n：

$$\begin{aligned}V_n&=\frac{1}{2}(V_1+V_o)\\&=\frac{1}{2}(V_o-2V_S+V_o)\\&=V_o-V_S\end{aligned}$$

最后，将 V_1 和 V_n 代入节点 1 的 KCL 方程，可得

$$-R_1V_n+(2R_1+R_2)V_1-R_1V_o=0$$

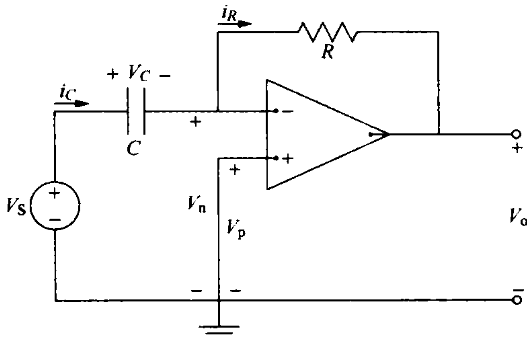
$$-R_1(V_o-V_S)+(2R_1+R_2)(V_o-2V_S)-R_1V_o=0$$

简化之后，有

$$V_o = \frac{(3R_1 + 2R_2)}{R_2} V_S$$ ■

下面两个例题说明如何利用电容构成微分运算和积分运算的运放电路。

例 8.23 请计算以下电路的输出电压 V_o。



解：由于运放同相输入端与地相连，因此，$V_p=0=V_n$。由 KVL 可得

$$V_C = V_S$$

于是

$$i_C = C\frac{dV_C}{dt} = C\frac{dV_S}{dt}$$

没有电流流入运放，因此，$i_C=i_R$。并且

$$i_R = \frac{V_n - V_o}{R} = -\frac{V_o}{R}$$

结合

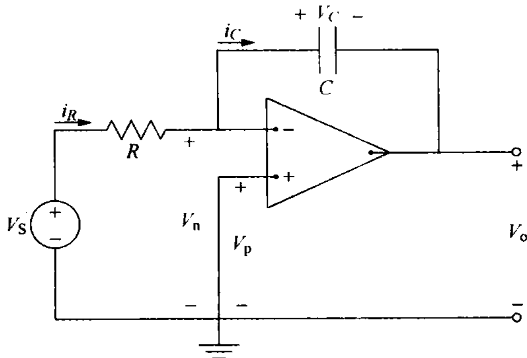
$$i_C = C\frac{dV_S}{dt} = i_R = -\frac{V_o}{R}$$

可得

$$V_o = -RC\frac{dV_S}{dt}$$

如果 $R=\frac{1}{C}$，本电路的输出就是输入的微分，即 $V_o=-\frac{dV_S}{dt}$。 ■

例 8.24 请计算以下电路的输出电压 V_o。



解：由前述可知，

$$V_n=V_p=0$$

并且

$$i_C=i_R=\frac{V_S}{R}$$

因此，有

$$V_C=\frac{1}{C}\int_{-\infty}^{t} i_C d\lambda=\frac{1}{C}\int_{-\infty}^{t}\frac{V_S}{R}d\lambda$$

由 KVL 可得

$$V_C+V_o=0$$

和

$$V_o=-\frac{1}{RC}\int_{-\infty}^{t} V_S d\lambda$$

如果 $R=\frac{1}{C}$，则本电路就是一个积分器，即

$$V_o=-\int_{-\infty}^{t} V_S d\lambda$$

■

运算放大器的电压特性

上述几个介绍运放应用的例题中都省略了图 8.29 所示的电源电压，实际上理想运放的输出电压受到电源电压的限制，只能处于 V^- 与 V^+ 之间。如果计算出来的输出电压 V_o 大于 V^+，那么，V_o 就会在 V^+ 处饱和；如果计算出来的 V_o 小于 V^-，那么，V_o 在 V^- 处饱和。如图 8.32 所示是运放的输出电压特性曲线。

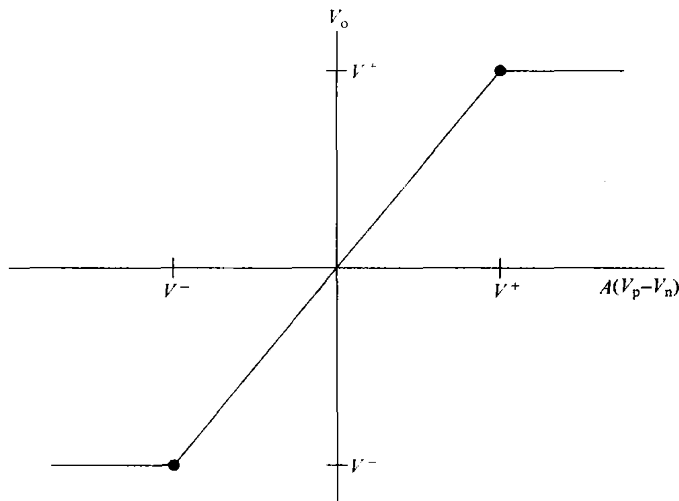


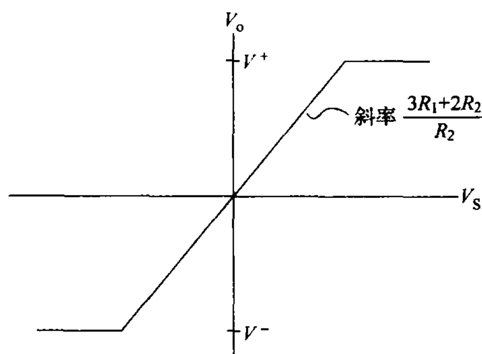
图 8.32　运放的电压特性

例 8.25　对于例 8.22 所示的电路，如果 $V^+=+10\text{V}$，$V^-=-10\text{V}$。请作出电路的输出电压特性曲线。

解：例 8.22 的解是

$$V_o=\left(\frac{3R_1+2R_2}{R_2}\right)V_S$$

其输出电压特性曲线如下，当 V_o小于 V^- 或者大于 V^+ 时输出饱和。



■

8.12 时变信号

一般仪器设备都用 50Hz 或者 60Hz 的正弦波交流电源供电，本章后面所述内容多数涉及正弦波电源的暂态响应，现在要介绍稳态响应，即强迫响应，它是生物医学仪器的实际运行状态。只分析稳态响应可以简化仪器的设计。正弦波电压源是一个时变信号，即

$$V_S=V_m\cos(\omega t+\phi) \tag{8.30}$$

式中 ω——角频率，单位是弧度/秒，即 rad/s；

ϕ——相角，单位是弧度或者度；

V_m——振幅。

正弦波周期 T、频率 f（单位是 Hz）和角频率 ω 之间的关系为

$$\omega=2\pi f=\frac{2\pi}{T} \tag{8.31}$$

方均根（root mean square，rms）是正弦波的一个重要指标，其计算公式为

$$V_{rms}=\sqrt{\frac{1}{T}\int_0^T V_m^2\cos^2(\omega t+\phi)\mathrm{d}t} \tag{8.32}$$

可以简化为 $V_{rms}=\dfrac{V_m}{\sqrt{2}}$。

为了考察时变输入 $V_S=V_m\cos(\omega t+\phi)$ 的输出响应，如图 8.33 所示，假设在 $t=0$ 时刻电路的开关闭合，并且电感中没有初始储能。那么，对该电路应用 KVL，有

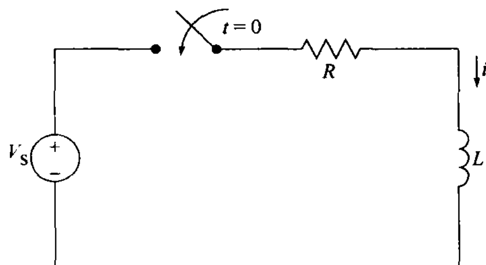


图 8.33 包含正弦波输入信号的 RL 电路

$$L\frac{\mathrm{d}i}{\mathrm{d}t}+iR=V_m\cos(\omega t+\phi)$$

求解此方程可得

$$\begin{aligned} i &= i_n + i_f \\ &= \frac{-V_m}{\sqrt{R^2+\omega^2L^2}}\cos\left(\phi-\frac{\omega L}{R}\right)e^{-\frac{R}{L}t}+\frac{V_m}{\sqrt{R^2+\omega^2L^2}}\cos\left(\omega t+\phi-\frac{\omega L}{R}\right) \end{aligned}$$

此式第一项 i_n是自由响应，当 t 趋于无穷大时，该项趋于 0；第二项 i_f是强迫响应，它与输入信号具有同样的形式，也是一个具有同样的角频率 ω 的正弦波，但是其相角和振幅与输入信号不同。生物医学仪器应用中一般只需要求解电路的稳态响应，在这里也就是只需求得幅值和相角两个未知量。下面介绍利用相量（phasor）求解这两个未知量的方法。

8.12.1 相量

相量是一个复数，它包括正弦波的振幅和相角，式（8.30）的正弦电压信号用相量表示就是

$$V=V_me^{j\phi}=V_m\angle\phi \tag{8.33}$$

其中，指数中的相角单位是弧度，而符号$\angle\phi$中的相角单位是度。在时域和相量域之间进行信号变换时，要用到复数代数，也要用到相量的三角函数形式，即

$$V=V_m(\cos\phi+j\sin\phi) \tag{8.34}$$

8.12.2 无源电路元件的相量表示法

要使用相量法求无源电路元件的稳态解，首先必须建立电阻、电感和电容这些无源元件的电压和电流的相量表示法。假设正弦波电流为

$$i=I_m\cos(\omega t+\theta)$$

其相量表示为

$$I=I_m\angle\theta=I_me^{j\theta}$$

对于电阻，其电压就是

$$v=IR=RI_m\cos(\omega t+\theta)$$

v 的相量表示就是

$$V=RI_m\angle\theta=RI \tag{8.35}$$

注意，电阻的相量电流和相量电压之间没有相移。

对于电感，则有

$$v=L\frac{\mathrm{d}i}{\mathrm{d}t}=-\omega LI_m\sin(\omega t+\theta)=-\omega LI_m\cos(\omega t+\theta-90^\circ)$$

该电压 V 的相量就是

$$
\begin{aligned}
V &= -\omega L I_m \angle\theta - 90^\circ = -\omega L I_m e^{j(\theta-90^\circ)} \\
&= -\omega L I_m e^{j\theta} e^{-j90^\circ} = -\omega L I_m e^{j\theta}(-j) \\
&= j\omega L I_m e^{j\theta} \\
&= j\omega L I
\end{aligned} \tag{8.36}
$$

注意，电感的电流与电压之间有 90°的相移，即电流滞后电压 90°。

对于电容，假设正弦波电压为 $v=V_m\cos(\omega t+\theta)$，其相量为 $V=V_m\angle\theta$，则电流为

$$
\begin{aligned}
i &= C\frac{dv}{dt} = C\frac{d}{dt}(V_m\cos(\omega t+\theta)) \\
&= -CV_m\omega\sin(\omega t+\theta) = -CV_m\omega\cos(\omega t+\theta-90^\circ)
\end{aligned}
$$

该电流 i 的相量就是

$$
\begin{aligned}
I &= -\omega CV_m \angle\theta - 90^\circ = -\omega CV_m e^{j\theta} e^{-j90^\circ} \\
&= -\omega CV_m e^{j\theta}\ (\cos\ (90^\circ)\ - j\sin\ (90^\circ)) \\
&= j\omega CV_m e^{j\theta} \\
&= j\omega CV
\end{aligned}
$$

于是，电压与电流关系的相量表示为

$$
V = \frac{1}{j\omega C} I = \frac{-j}{\omega C} I \tag{8.37}
$$

注意，电容的电流与电压之间也有 90°的相移，是电压滞后电流 90°。

式（8.35）～式（8.37）都有 $V=ZI$ 的形式，其中 Z 代表电路元件的阻抗，通常是一个复数，单位为 Ω。电阻的阻抗是 R，电感的阻抗是 $j\omega L$，电容的阻抗则是 $\frac{-j}{\omega C}$。阻抗是个复数，它不是相量，虽然可能看起来像相量。阻抗的虚部称为电抗。

现在可以将整个电路图从时域转换到相量域了。例如，图 8.34 所示的电路转换到相量域就是图 8.35，也就是把每个无源元件用阻抗表示，电源用相量表示。其中，电压源为 $V_S = 100\sin500t = 100\cos(500t-90^\circ)\text{mV} \leftrightarrow 500\angle-90^\circ$。

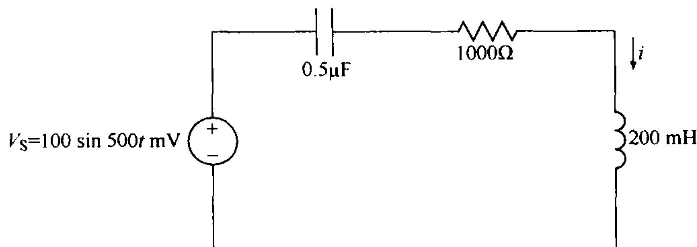


图 8.34 电路图

电容为

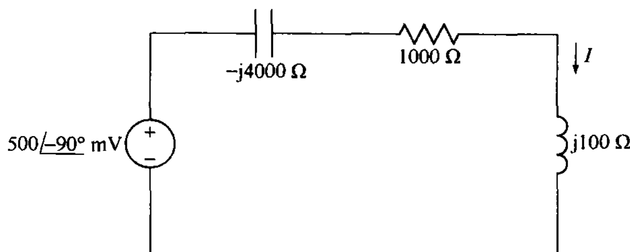


图 8.35 图 8.34 的相量和阻抗等效电路

$$0.5\mu F \leftrightarrow \frac{-j}{\omega C} = -j4000\Omega$$

电阻为

$$1000\Omega \leftrightarrow 1000\Omega$$

电感为

$$200mH \leftrightarrow j\omega L = j100\Omega$$

将每个元件用相量和阻抗等效就形成了图 8.35 所示的相量电路。

8.12.3 基尔霍夫定律等电路分析法的相量形式

所有本章前面所述的基尔霍夫电流定律和电压定律以及其他方法都可以用于相量域。在相量域中，KVL 就是任何闭合回路的相量电压之和为 0，即

$$\sum V_i = 0 \tag{8.38}$$

KCL 就是离开任何节点的电流之和为 0，即

$$\sum I = 0 \tag{8.39}$$

串联阻抗为

$$Z = Z_1 + \cdots + Z_n \tag{8.40}$$

并联阻抗为

$$Z = \frac{1}{\frac{1}{Z_1} + \cdots + \frac{1}{Z_n}} \tag{8.41}$$

节点电压法、叠加原理以及戴维南等效电路也都可以用于相量域。下面两个例题说明了这些方法的应用，最难的是用到了复数代数。

例 8.26 请计算图 8.35 所示电路的稳态响应电流 i。

解： 该电路的阻抗为

$$Z = -j4000 + 1000 + j100 = 1000 - j3900\Omega$$

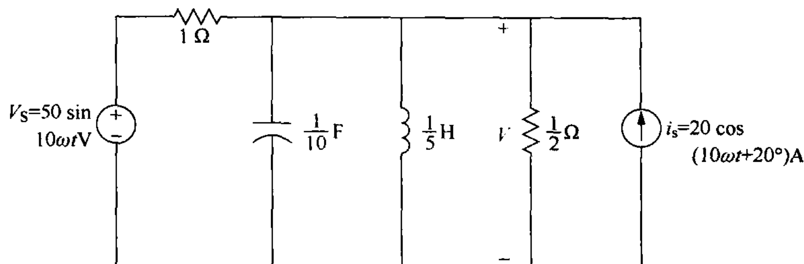
应用欧姆定律，可得

$$I = \frac{V}{Z} = \frac{0.5\angle -90^\circ}{1000 - j3900} = \frac{0.5\angle -90^\circ}{4026\angle -76^\circ} = 124\angle -14^\circ\ \mu A$$

将其变换到时域，可得稳态电流为

$$i=124\cos(500t-14°)\mu A$$

例 8.27 请利用节点电压法计算以下电路的稳态响应电压 v。



解： 第一步是将电路元件转化为相量阻抗。其中，电容和电感的阻抗分别为

$$\frac{1}{10}F\leftrightarrow\frac{-j}{\omega C}=-j\Omega$$

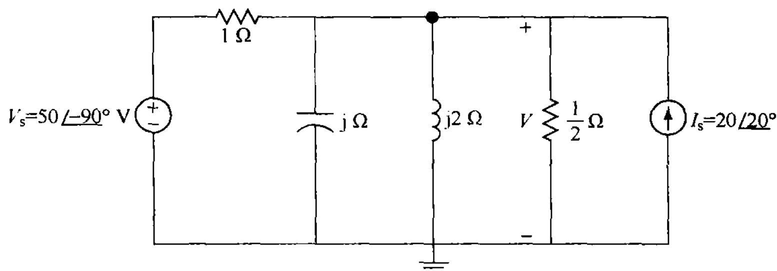
$$\frac{1}{5}H\leftrightarrow j\omega L=j2\Omega$$

电压源 V_S和电流源 i_S的相量表示为

$$v_S=50\sin\omega tV\leftrightarrow V_S=50\angle-90°V$$

$$i_S=20\cos(\omega t+20°)\ A\leftrightarrow I_S=20\angle 20°$$

两个电阻保持原值。于是，电路的相量图如下所示，电路下方节点为地。



写出节点 1 的节点电压方程，即

$$V-50\angle-90°+\frac{V}{-j}+\frac{V}{j2}+2V-20\angle 20°=0$$

合并其同类项，转化成直角形式并转化成极坐标形式，有

$$V\left(3+\frac{j}{2}\right)=50\angle-90°+20\angle 20°$$

$$V\left(3+\frac{j}{2}\right)=-50j+18.8+j6.8=18.8-j43.2$$

$$V\times 3.04\angle 9.5°=47.1\angle-66.5°$$

$$V=\frac{47.1\angle -66.5^\circ}{3.04\angle 9.5^\circ}=15.5\angle -76^\circ$$

于是，电压 v 的稳态解就是

$$v=15.6\cos(10t-76^\circ)\text{V}$$

■

8.13　有源模拟滤波器

本节将介绍几种运算放大器构成的有源模拟滤波器。只用电阻、电容和电感这些无源元件构成的滤波器称为无源滤波器，用运放构成的滤波器则称为有源滤波器。要改善无源模拟滤波器的性能，通常需要在滤波器的输出端增加负载阻抗。但是，如果用了运放，那么，不增加滤波器输出端的负载也可以很好地调控滤波器的性能。滤波器用于去除测量信号中所含的噪声。根据信号的频域特性设计滤波器，可以去除噪声，保留所需的被测信号。

图 8.36 所示是理想低通、高通、带通和陷波 4 种滤波器的频域特性，其中，信号可以通过的频率范围称为通带，被滤波器去除的频率范围称为阻带。滤波器的幅值（即 $|H(j\omega)|$）在通带处为 1，在阻带处为 0。低通滤波器允许频率小于 ω_1 的缓慢变化的信号通过滤波器，而任何频率大于 ω_1 的信号和噪声则被去除。高通滤波器允许频率大于 ω_2 的快速变化信号通过滤波器，任何频率小于 ω_2 的信号和噪声则被去除。带通滤波器则允许频率大于 ω_1 且小于 ω_2 的信号通过，此范围之外任何的信号和噪声被去除。陷波滤波器则允许频率小于 ω_1 或者大于 ω_2 的信号通过，其他频率的信号和噪声被去除。频率 ω_1 和 ω_2 通常称为滤波器的截止频率。

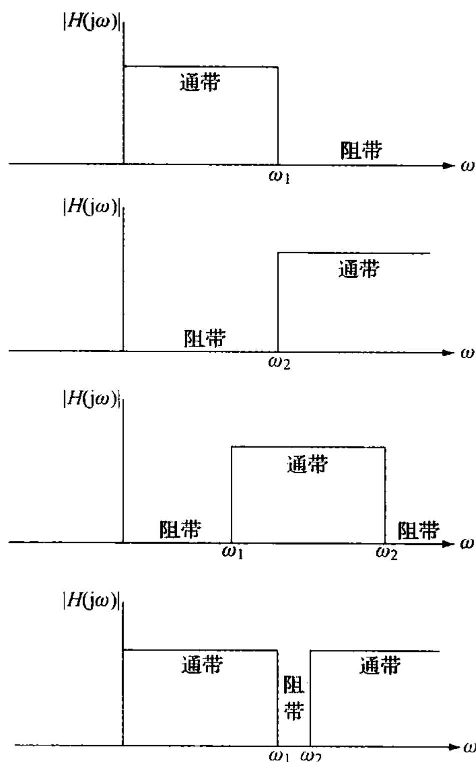


图 8.36　4 种理想滤波器的幅频特性

注：从上到下为低通、高通、带通和陷波滤波器。

但是，实际使用的滤波器不可能有这样的理想特性，如图 8.37 所示，这些滤波器在通带到阻带之间都有过渡阶段，并不是突变的，第 10 章会介绍其中的原因。另外，用同一个电路同时实现放大和滤波两种功能是很方便的，因此，滤波器的最大

幅值并不一定是 1，可以根据需要设定为任意值 M。

测定滤波器的性能时，在滤波器输入端接入正弦波信号，并在整个需要测试的频率范围内，逐一改变正弦波频率，同时测量各个离散频率点上的滤波器输出幅值。$|H(\mathrm{j}\omega)|=\frac{M}{\sqrt{2}}$时的频率即为截止频率。

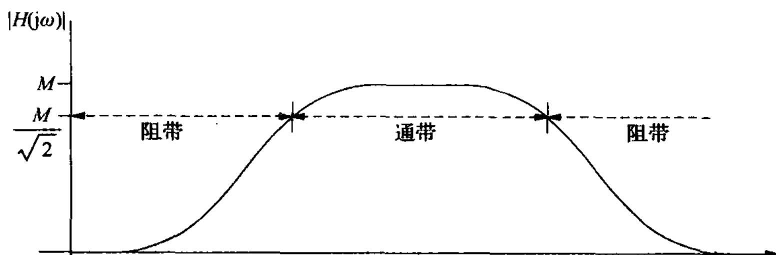
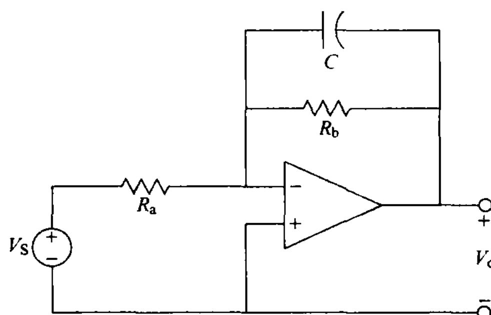


图 8.37　实际带通滤波器的幅频特性

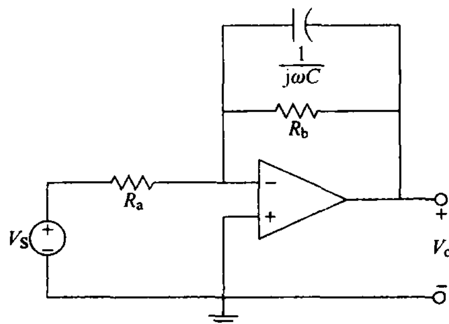
注：最大幅值 M 并不一定为 1，通带定义为幅值大于$\frac{M}{\sqrt{2}}$的频率范围。

例 8.28　请将下图所示低通滤波器电路的增益设计为 5，截止频率设计为 500rad/s。



解：将电路中的运放看作理想的，注意，其同相端接地，所以，反相端也相当于接地。可见该滤波器的工作原理是：对于低频信号，电容相当于开路，电路简化为反相放大器，可以通过低频信号；对于高频信号，电容相当于短路，将运放输出端与反相输入端相连，也就是与地相连，因此，不能通过高频信号。

如下图所示，先将电路转化到相量域，再用相量法求解。



离开反相输入端节点的电流之和为

$$-\frac{V_S}{R_a}-\frac{V_o}{\frac{1}{j\omega C}}-\frac{V_o}{R_b}=0$$

重排该方程并合并同类项，可得

$$-V_o\left(\frac{1}{\frac{1}{j\omega C}}+\frac{1}{R_b}\right)=\frac{V_S}{R_a}$$

也就是

$$\frac{V_o}{V_S}=-\frac{1}{R_a}\left(\frac{1}{\frac{1}{\frac{1}{j\omega C}}+\frac{1}{R_b}}\right)=-\frac{1}{R_a}\left(\frac{1}{j\omega C+\frac{1}{R_b}}\right)$$

$$\frac{V_o}{V_S}=-\frac{1}{R_aC}\left(\frac{1}{j\omega+\frac{1}{R_bC}}\right)$$

截止频率定义为 $\omega_c=\frac{1}{R_bC}$，因此，要使 $\omega_c=500\ \frac{\text{rad}}{\text{s}}$，那么$\frac{1}{R_bC}=500$。另一方面，截止频率也定义为$|H(j\omega)|=\frac{M}{\sqrt{2}}$，本题要求 $M=5$，而 V_o/V_S的幅值为

$$\left|\frac{V_o}{V_S}\right|=\frac{\frac{1}{R_aC}}{\sqrt{\omega^2+\left(\frac{1}{R_bC}\right)^2}}$$

在截止频率处有

$$\frac{5}{\sqrt{2}}=\frac{\frac{1}{R_aC}}{\sqrt{\omega_c^2+\left(\frac{1}{R_bC}\right)^2}}$$

代入 $\omega_c=500\ \frac{\text{rad}}{\text{s}}$和$\frac{1}{R_bC}=500$，该式变为

$$\frac{5}{\sqrt{2}}=\frac{\frac{1}{R_aC}}{\sqrt{\omega_c^2+\left(\frac{1}{R_bC}\right)^2}}=\frac{\frac{1}{R_aC}}{\sqrt{500^2+500^2}}=\frac{\frac{1}{R_aC}}{500\sqrt{2}}$$

由此可得

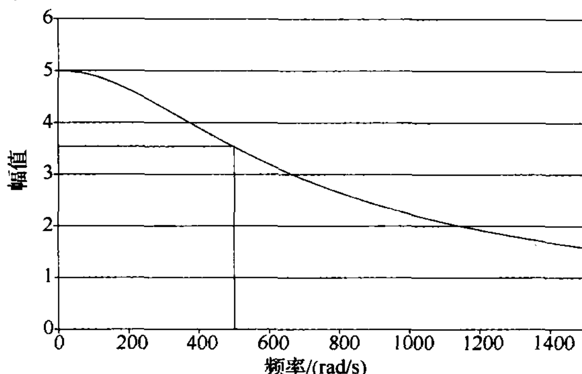
$$R_aC=\frac{1}{2500}$$

由于两个方程（$R_aC=\frac{1}{2500}$和$\frac{1}{R_bC}=500$）中包含了 3 个未知量，因此，本题有无穷多个解。可以选择便利的元件值，例如选 $R_a=20\text{k}\Omega$，则其他两个元件的值就是

$$C=\frac{1}{2500\times R_a}=\frac{1}{2500\times 20000}=20\text{nF}$$

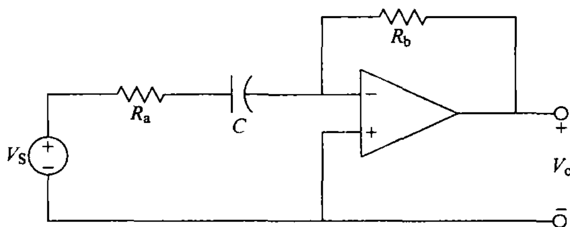
$$R_b=\frac{1}{500\times C}=\frac{1}{500\times 20\times 10^{-9}}=100\text{k}\Omega$$

下图显示了该电路的幅值随频率变化的曲线。可见，截止频率 500 $\frac{\text{rad}}{\text{s}}$处的幅值等于 3.53，符合设计要求。

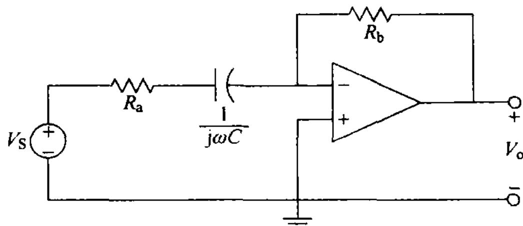


■

例 8.29 请将下图所示高通滤波器电路的增益设计为 5，截止频率设计为 100rad/s。



解：假设运放是理想的，由于其同相端接地，反相端也相当于接地。可见该滤波器的工作原理是：对于低频信号，电容相当于开路，输入电压与反相输入端相当于断开，没有输入，输出也就为 0；对于高频信号，电容相当于短路，于是电路简化为反相放大器，因此高频信号可以通过。与前面的例 8.28 相似，如下图所示，先将电路转化到相量域，再用相量法求解。



此电路中，离开反相输入端节点的电流之和为

$$-\frac{V_S}{R_a+\frac{1}{j\omega C}}-\frac{V_o}{R_b}=0$$

重排方程可得

$$\frac{V_o}{V_S}=-\frac{R_b}{R_a+\frac{1}{j\omega C}}=-\frac{R_b}{R_a}\frac{j\omega}{j\omega+\frac{1}{R_aC}}$$

$\frac{V_o}{V_S}$的幅值为

$$\left|\frac{V_o}{V_S}\right|=\frac{R_b}{R_a}\frac{\omega}{\sqrt{\omega^2+\left(\frac{1}{R_aC}\right)^2}}$$

在截止频率 $\omega_c=100\ \frac{\text{rad}}{\text{s}}=\frac{1}{R_aC}$处，有

$$\frac{5}{\sqrt{2}}=\frac{R_b}{R_a}\frac{\omega_c}{\sqrt{\omega_c^2+\left(\frac{1}{R_aC}\right)^2}}$$

将$\frac{1}{R_aC}=100$ 和 $\omega_c=100\ \frac{\text{rad}}{\text{s}}$代入，可得

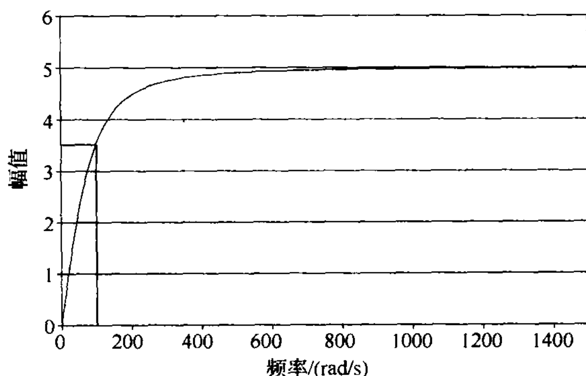
$$\frac{5}{\sqrt{2}}=\frac{R_b}{R_a}\frac{\omega_c}{\sqrt{\omega_c^2+\left(\frac{1}{R_aC}\right)^2}}=\frac{R_b}{R_a}\frac{\frac{1}{R_aC}}{\sqrt{100^2+100^2}}=\frac{R_b}{R_a}\frac{100}{100}=\frac{R_b}{\sqrt{2}R_a}$$

于是$\frac{R_b}{R_a}=5$。由于两个方程中包含了三个未知量，因此，如果选定其中一个元件的值，例如 $R_b=20\text{k}\Omega$，就可以计算其他两个元件的值，即

$$R_a=\frac{R_b}{5}=\frac{20000}{5}=4\text{k}\Omega$$

$$C=\frac{1}{100R_a}=\frac{1}{100\times4000}=2.5\mu\text{F}$$

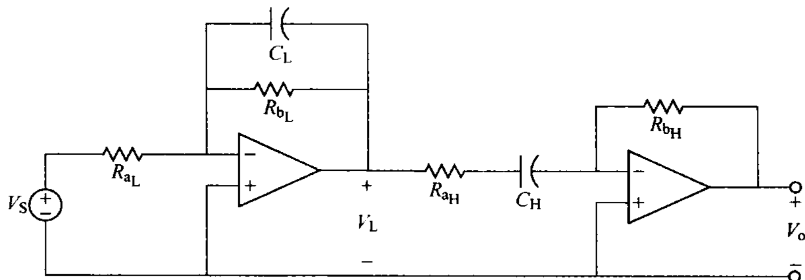
下图显示了此电路的幅值特性随频率变化的曲线。可见，截止频率 100 $\frac{\text{rad}}{\text{s}}$处的幅值等于 3.53，符合设计要求。



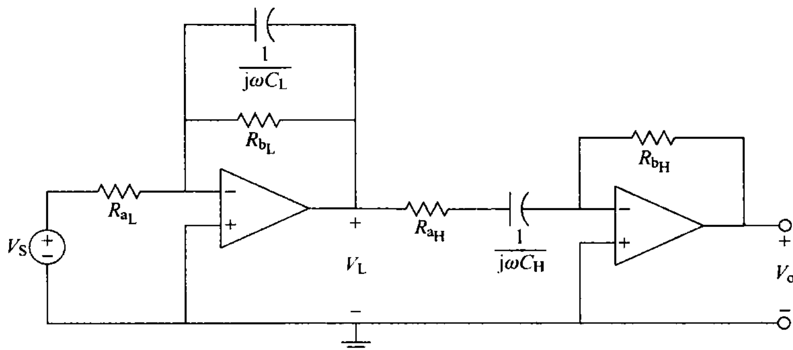
■

下面举例说明具有两个截止频率的带通滤波器的设计方法。

例 8.30 请将下图所示带通滤波器电路的增益设计为 5，通带频率范围设计为 100～500rad/s。



解：该滤波器的设计也在相量域进行，并且可以利用前两个例题的计算结果。注意，左边的运放组成了一个低通滤波器电路，右边的运放则组成了一个高通滤波器，两个滤波器级联形成一个带通滤波器。实际上，通过运放的级联可以设计各种不同类型的滤波器。下图是其相量域的电路图。



与前两个例题一样，本电路两个运放的同相端都接地，这意味着两个反相端也相当于接地。离开这两个反相输入端节点的电流之和分别为

$$-\frac{V_S}{R_{a_L}}-\frac{V_L}{\frac{1}{j\omega C_L}}-\frac{V_L}{R_{b_L}}=0$$

和

$$-\frac{V_L}{R_{a_H}+\frac{1}{j\omega C_H}}-\frac{V_o}{R_{b_H}}=0$$

求解第一个方程，可得 V_L 为

$$V_L=-\frac{1}{R_{a_L}C_L}\left(\frac{1}{j\omega+\frac{1}{R_{b_L}C_L}}\right)V_S$$

求解第二个方程，可得 V_o 为

$$V_o=-\frac{R_{b_H}}{R_{a_H}}\frac{j\omega}{j\omega+\frac{1}{R_{a_H}C_H}}V_L$$

将 V_L代入此方程，可得

$$V_o=\frac{R_{b_H}}{R_{a_H}}\frac{j\omega}{j\omega+\frac{1}{R_{a_H}C_H}}\times\frac{1}{R_{a_L}C_L}\left(\frac{1}{j\omega+\frac{1}{R_{b_L}C_L}}\right)V_S$$

这个解其实是两个滤波器各自解的乘积。滤波器的幅值为

$$\left|\frac{V_o}{V_S}\right|=\frac{R_{b_H}}{R_{a_H}}\frac{\omega}{\sqrt{\omega^2+\left(\frac{1}{R_{a_H}C_H}\right)^2}}\frac{\frac{1}{R_{a_L}C_L}}{\sqrt{\omega^2+\left(\frac{1}{R_{b_L}C_L}\right)^2}}$$

根据给定的两个截止频率可以产生如下两个方程

$$\omega_{c_H}=\frac{1}{R_{a_H}C_H}=100\ \frac{\text{rad}}{\text{s}}$$

和

$$\omega_{c_L}=\frac{1}{R_{b_L}C_L}=500\ \frac{\text{rad}}{\text{s}}$$

每个截止频率处幅值都是$\frac{5}{\sqrt{2}}$，因此，在 $\omega_{c_H}=100\ \frac{\text{rad}}{\text{s}}$处有

$$\frac{5}{\sqrt{2}}=\frac{R_{b_H}}{R_{a_H}}\frac{\omega_{c_H}}{\sqrt{\omega_{c_H}^2+\left(\frac{1}{R_{a_H}C_H}\right)^2}}\frac{\frac{1}{R_{a_L}C_L}}{\sqrt{\omega_{c_H}^2+\left(\frac{1}{R_{b_L}C_L}\right)^2}}$$

$$=\frac{R_{b_H}}{R_{a_H}}\frac{100}{\sqrt{100^2+100^2}}\frac{\frac{1}{R_{a_L}C_L}}{\sqrt{100^2+500^2}}$$

也就是

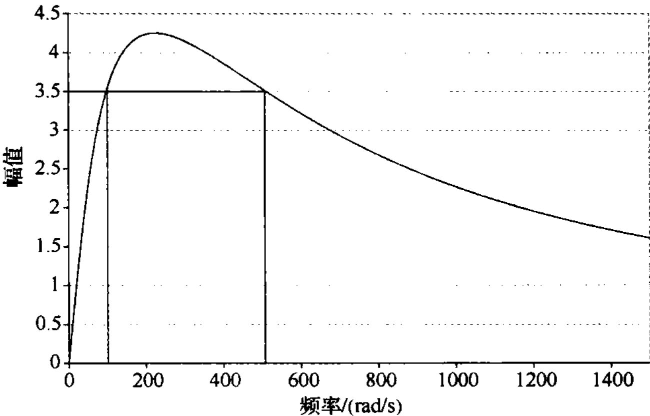
$$500\sqrt{26}=\frac{R_{b_H}}{R_{a_H}R_{a_L}C_L}$$

由另一个截止频率可以得到同样的方程。现在共有如下 3 个方程：$\frac{1}{R_{a_H}C_H}=100$、$\frac{1}{R_{b_L}C_L}=500$ 和 $500\sqrt{26}=\frac{R_{b_H}}{R_{a_H}R_{a_L}C_L}$；但有 6 个未知量。如果选定 $R_{b_L}=100\text{k}\Omega$ 且 $R_{a_H}=100\text{k}\Omega$，就可以求得 $C_L=\frac{1}{500R_{b_L}}=20\text{nF}$ 和 $C_H=\frac{1}{100R_{a_H}}=0.1\mu\text{F}$。然后，由 $500\sqrt{26}=\frac{R_{b_H}}{R_{a_H}}\frac{1}{R_{a_L}C_L}$可得

$$\frac{R_{b_H}}{R_{a_L}}=500\sqrt{26}C_LR_{a_H}=5.099$$

因此，如果再选定一个电阻值，例如 $R_{a_L}=10\text{k}\Omega$，就可得 $R_{b_H}=50.099\text{k}\Omega$。

下图显示了此电路的幅频特性曲线。可见，截止频率 100rad/s 和 500rad/s 处的幅值都等于 3.53，符合设计要求。



■

由此可见，例 8.28～例 8.30 中没有一个滤波器具有图 8.36 所示的理想特性。如果将同样的滤波器级联起来，也就是把前一个滤波器的输出接到后一个滤波器的输入上，以此类推，可以使低通滤波器通带到阻带的过渡阶段变得较为陡峭，改善滤波器的特性。级联的滤波器越多，过渡带特性就越好。这时，整个级联滤波器的幅值就等于各个滤波器幅值之积。

虽然这种方法可以有效地改善滤波器过渡带的特性，但是，这种滤波器通带内的幅值不再是恒定不变的常数。除了这种简单的级联滤波器以外，还有性能更好的滤波器，例如巴特沃思滤波器（Butterworth filter）。图 8.38 显示了两个巴特沃思滤

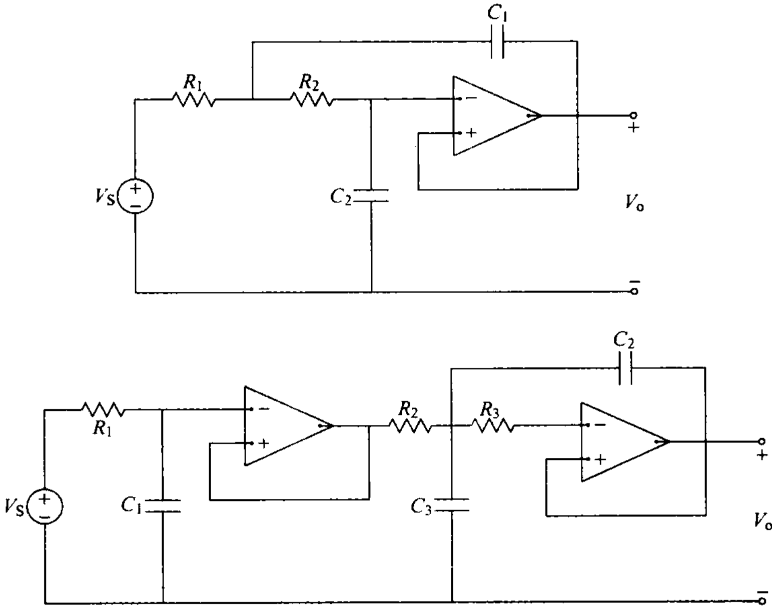


图 8.38 上图和下图分别是二阶和三阶巴特沃思低通滤波器

波器，有关这两个滤波器的分析计算作为习题 54 和 55 留给读者。

8.14 生物医学仪器设计

图 8.2 显示了生物医学仪器系统的各个主要组成部分，这类仪器的用途是监测传感器的输出信号并且提取其中的有用信息，本书第 9 章将介绍生物医学传感器，第 10 章将介绍测量信号的处理方法。

仪器系统中的模/数转换器（Analog-to-Digital Converter，A/D）将连续时间信号转化成离散时间信号，并将数据存储到计算机内存中，也就是将连续时间波形采样成为每隔 t_k 秒取一个的数据序列，同时也将波形的模拟量变成数字量，即将数据的幅值变成 2^n 个离散值之一⊖，这样，连续时间波形就转化成了一个个计算机的字节，存放在内存中。要完整地采集连续时间信号，就必须谨慎选择采样时间间隔 t_k，以免丢失信息。第 10 章将会介绍，根据通信理论中的采样定理，最小采样频率必须是信号所含最高频率的 2 倍。实际应用中，经常用信号最高频率的 5～10 倍作为采样频率，以便减少混叠引起的误差，获得较高的采样精度。

8.14.1 噪声

生物医学系统的测量信号中不可避免地总是含有噪声，电源线路、无线电发射信号和电视机的电磁波等噪声源都会带来干扰噪声。通过仔细设计电路的连接结构，减小耦合效应，可以有效地减少噪声干扰。

噪声的来源有很多，电源线（50Hz 或 60Hz）、荧光灯、AM/FM 无线电广播、计算机时钟的振荡器、实验室的各种设备、手机等都会引入干扰噪声。噪声源的电磁辐射通过电容和电感的耦合进入放大器电路或者进入人体，甚至人体神经纤维上传导的动作电位也会在传感器和放大器的接口处产生噪声。在 A/D 转换器的输入端使用滤波器可以减少这些噪声，提高信噪比（Signal-to-Noise Ratio，SNR）。

在滤波器设计中，放大器的直流偏置、传感器的漂移、温漂等低频噪声可以用截止频率高于噪声频率但低于生物信号频率的高通滤波器来消除。神经传导、无线电广播、计算机、手机等高频噪声则可以用截止频率低于噪声频率但高于生物信号频率的低通滤波器来消除。电源线噪声在生物信号监测中是一个令人头痛的问题，因为 50Hz 或 60Hz频率通常在所测生物信号的频率范围之内。一般用陷波滤波器来消除这种噪声，陷波频率设定为 50Hz 或者 60Hz，截止频率则位于陷波频率两边几赫兹处。

还有一类噪声被称为固有噪声（inherent noise），它来自电子元件的随机过程，这种随机过程是电子元件的工作基础。如果电路设计得比较好，可以减小这种噪声。但是，固有噪声只能减小，不可能彻底消除。低通滤波器可以减少噪声的高频成分，但是，在生物信号频率范围内的噪声成分就无法用滤波方法去除。

⊖ n 为数字量的二进制位数。——译者注

8.14.2 计算机

计算机有 3 个基本组成部件：中央处理器（Central Processing Unit，CPU）、算术逻辑运算器（Arithmetic and Logic Unit，ALU）和内存（Memory）。CPU 通过执行程序指令，指挥计算机中所有其他部件的运行，并且控制运行过程中各部件之间的信息流通；ALU 则完成加、减、乘、除等各种算术运算以及与、或、非等逻辑运算。

计算机内存有只读存储器（Read Only Memory，ROM）和随机存取存储器（Random Access Memory，RAM）两种。ROM 是一种永久性保存编程内容的集成电路，它存储 CPU 运行所需的最基本程序，用户不能改变其内容；RAM 则用于信息的暂时存储，用户可以改变其内容。用户编写的程序、输入数据、处理结果等都保存在 RAM 中。

计算机是使用二进制的设备，它用“1”表示高电平信号，用“0”表示低电平信号。电信号都组合成 8 位二进制数的形式，称为“字节”，用于编码各种信息。两个字节组合起来成为一个“字”。目前使用的一般台式计算机都是 32 位系统，也就是总共可以组合产生约 4.29×10^9 个地址编码，用于访问内存空间。最早的计算机只有 8 位，只能访问 256 个内存地址。

计算机程序设计语言将指令和数据与特定的一组二进制码对应起来，每种特定排列的编码只对应一种含义。字母和其他标点符号等也用特定的编码来表示，如 ASCII 码就是其中一种编码字符的方法，其全称是“美国标准信息交换码（American Standard Code for Information Exchange)”。ASCII 码提供了一种不同计算机类型之间信息交换的通用标准。将文字处理文件保存为文本文件（text）就是使用 ASCII 码格式。通常，不同的软件用不同的二进制格式保存文字处理文件，但是，几乎所有数据处理软件都可以导入或者导出 ASCII 码格式的数据文件。

最低级的计算机语言是机器语言，它由计算机可以理解的“0”和“1”组成，机器语言是计算机的自然语言。其上一级语言是汇编语言，它将英文缩写词与二进制码对应起来，以便于阅读。用汇编语言编写的程序可以直接操作内存空间，并且，这种程序的运行速度很快，常用于需要快速采样大量数据的采集系统。例如，用很高的采样频率从一组传感器采集数据就需要用汇编语言编写的程序来实现。

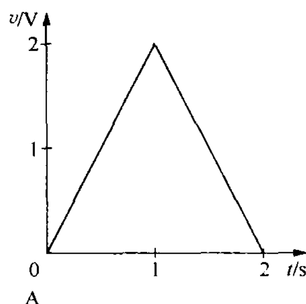
高级计算机语言如 FORTRAN、PERL 和 C++ 等由语句组成，每条语句的执行其实需要许多条机器语言指令或者汇编语言指令才能完成。高级语言的语句很像英语，还包含通用的数学符号。高级语言比机器语言和汇编语言容易学习和掌握。编写程序指令是为了告诉计算机何时并且如何利用各种硬件完成特定的工作，因此，这些指令必须按照正确的顺序传递给计算机的 CPU 才能获得预期的结果。

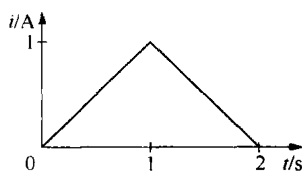
例如，使用计算机采集生理数据时，就必须编写程序，告诉计算机什么时候开

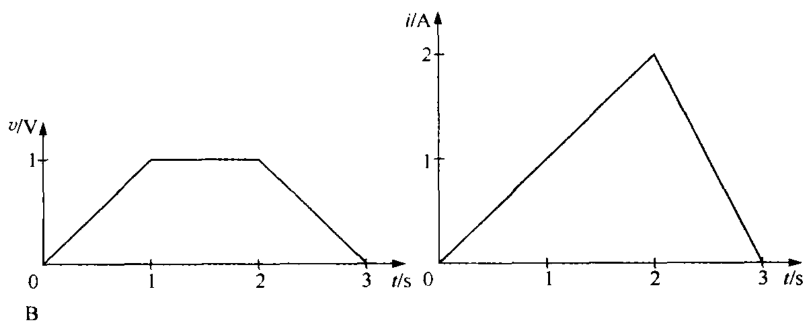
始采集数据、从哪些传感器以怎样的频率采集数据、需要采集多少数据量以及采集的数据存放在何处等信息。系统可以提供的最高采样频率决定于计算机时钟（如233MHz）限定的速度以及采样每个数据所需执行的计算机指令数。有些计算机还可以控制输入放大器的增益，在采样时自动调节信号的大小，而另一些系统的输入放大器增益只能手工调节。

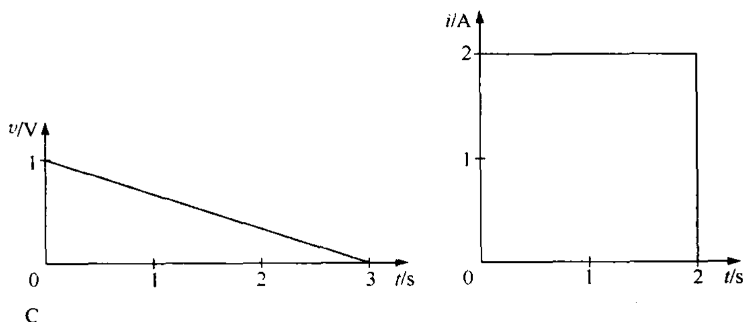
习　题

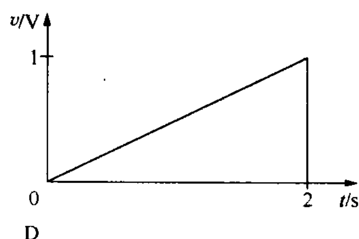
1. 假设电压 v 和电流 i 的波形如下，请分别计算 A～E 的各种情况下图 8.7 中电路元件所吸收的功率。

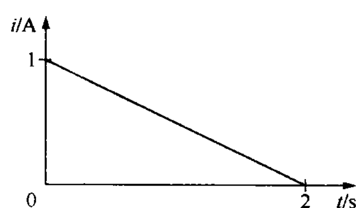


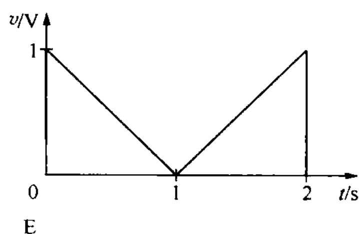


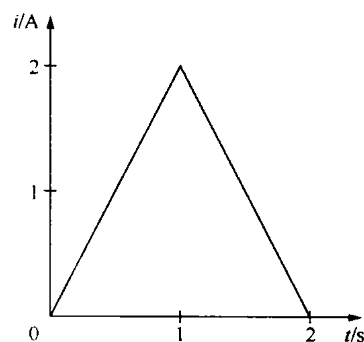












2. 假设图 8.7 电路两个端口之间的电压 v 和电流 i 如下

$$v = te^{-10000t}u(t)\,\text{V}$$

$$i = (t+10)e^{-10000t}u(t)\,\text{A}$$

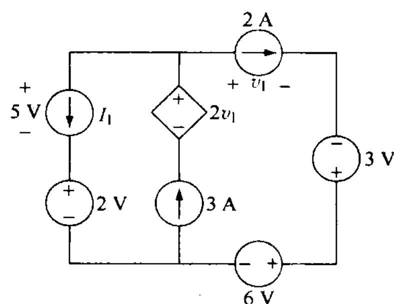
a. 求最大功率的时刻；

b. 求最大功率值；

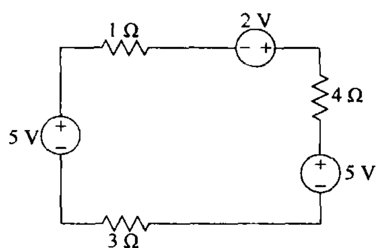
c. 计算 $t=1\times10^{-4}$ s 时电路接收的能量；

d. 计算电路接收到的总能量。

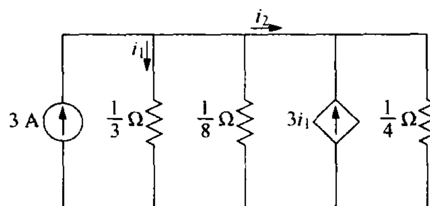
3. 试计算以下电路中的电压 v_1 以及各电路元件吸收或输送的电功率。



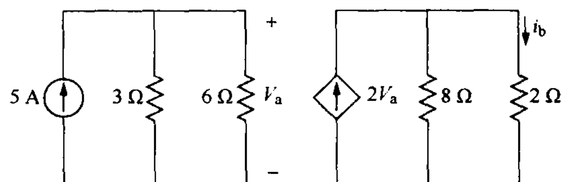
4. 试计算以下电路中每个元件的功率。



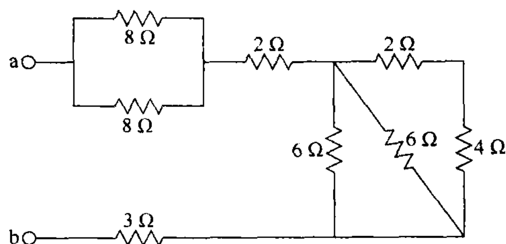
5. 试计算以下电路中的电流 i_2。



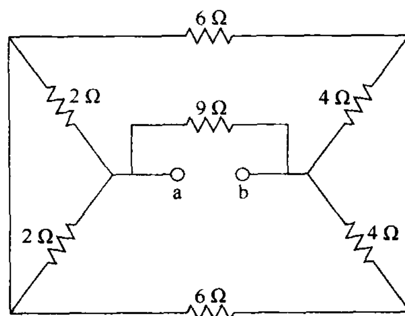
6. 试计算以下电路中电流 i_b。



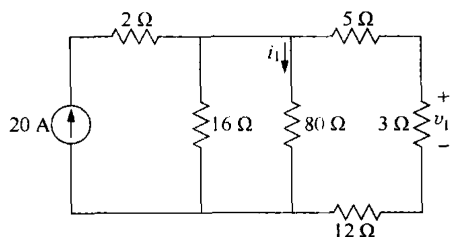
7. 试计算以下电路的等效阻抗 R_{ab}。



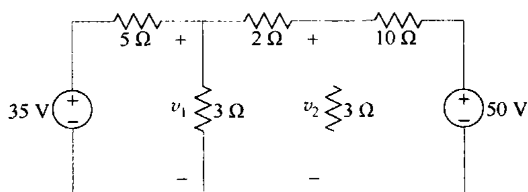
8. 试计算以下电路的等效阻抗 R_{ab}。



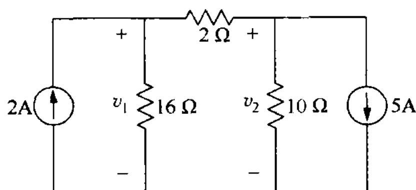
9. 试计算以下电路中的电流 i_1 和电压 v_1。



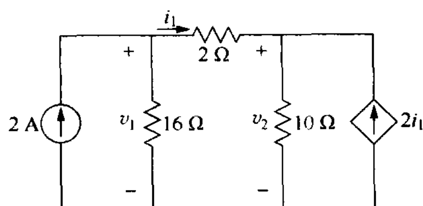
10. 请使用节点电压法计算以下电路中的电压 v_1 和 v_2。



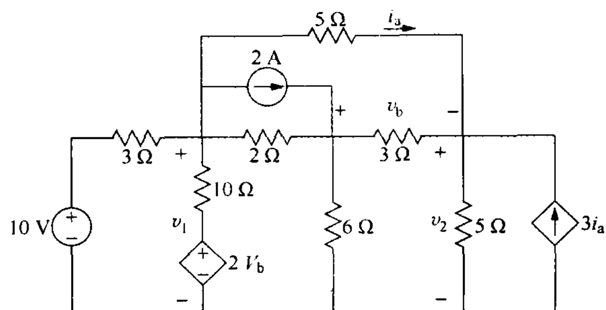
11. 请使用节点电压法计算以下电路中的电压 v_1 和 v_2。



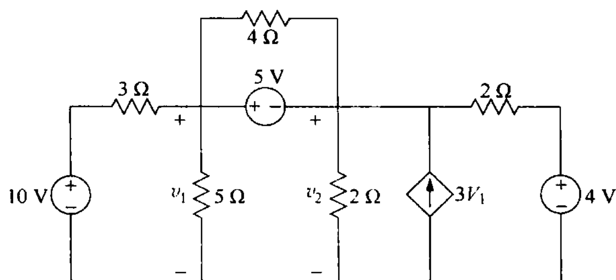
12. 请使用节点电压法计算以下电路中的电压 v_1 和 v_2。



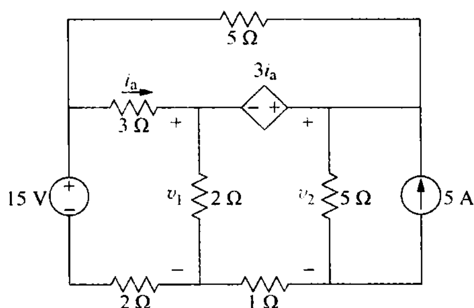
13. 请使用节点电压法计算以下电路中的电压 v_1 和 v_2。



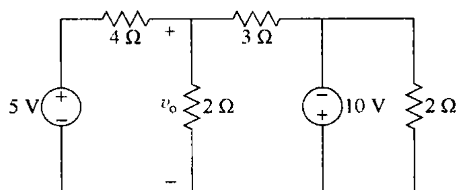
14. 请使用节点电压法计算以下电路中的电压 v_1 和 v_2。



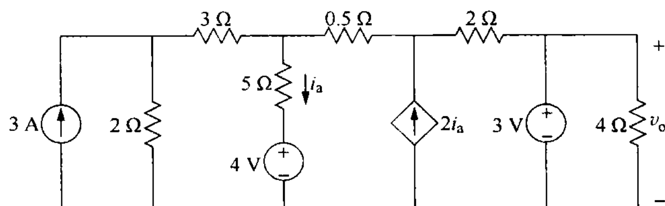
15. 请使用节点电压法计算以下电路中的电压 v_1 和 v_2。



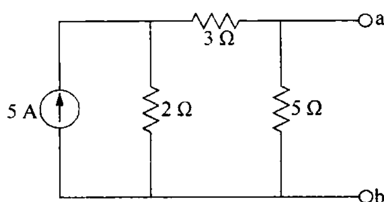
16. 请利用叠加原理计算以下电路中的电压 v_o。



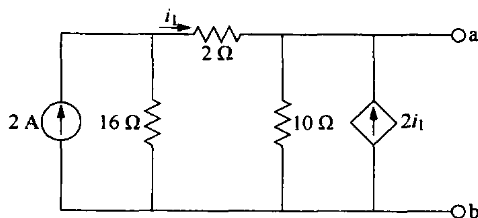
17. 请利用叠加原理计算以下电路中的电压 v_o。



18. 试求出以下电路 a 和 b 两端之间的戴维南等效电路。

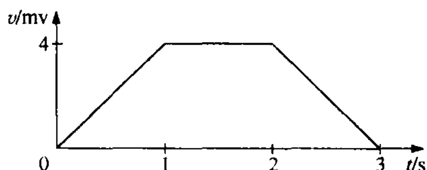


19. 试求出以下电路 a 和 b 两端之间的戴维南等效电路。



20. 某个 10mH 的电感上施加电流脉冲 $i(t)=(2+10e^{-2t})u(t)$，(a) 试计算电感两端的电压；(b) 请作出电流和电压曲线图；(c) 请计算功率的时间函数。

21. 某个电感两端的电压如下图所示，如果 $L=30\text{mH}$，$i(0)=0\text{A}$，试求 $t\geqslant0$ 时的电流变化曲线 $i(t)$。



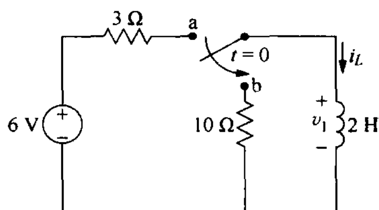
22. 某个 $4\mu F$ 电容两端的电压为 $v(t)=(200000t-50000)e^{-2000t}u(t)\text{V}$，试计算 (a) 电容的电流；(b) 功率的时间函数；(c) 电能量。

23. 某个 $5\mu\text{F}$ 电容上流过的电流为

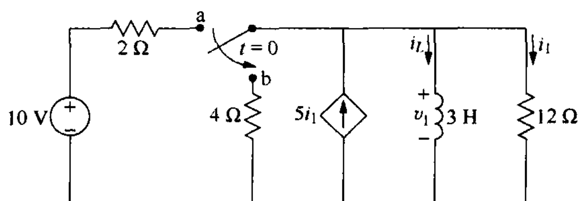
$$i(t)=\begin{cases}0\text{mA} & t<0\text{ms}\\ 5t^2\text{mA} & 0\leqslant t<1\text{ms}\\ 5(2-t^2)\text{mA} & t\geqslant1\text{ms}\end{cases}$$

请计算电容两端的电压。

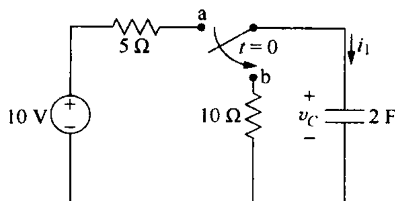
24. 如下图所示，假设开关已处于位置 a 足够长的时间。在 $t=0$ 时刻，开关被瞬时切换到位置 b，请计算 $t>0$ 时的电流 i_L 和电压 v_1。



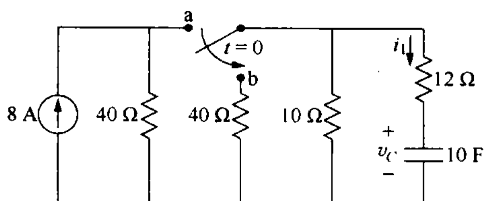
25. 如下图所示，假设开关已处于位置 a 足够长的时间。在 $t=0$ 时刻，开关被瞬时切换到位置 b，请计算 $t>0$ 时的电流 i_L、i_1 和电压 v_1。



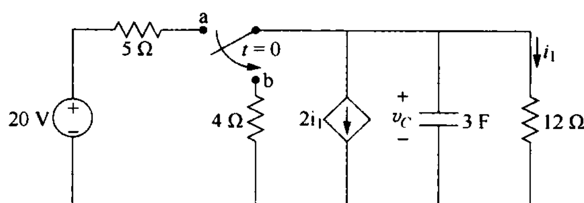
26. 如下图所示，假设开关已处于位置 a 足够长的时间。在 $t=0$ 时刻，开关被瞬时切换到位置 b，请计算 $t>0$ 时的电压 v_C 和电流 i_1。



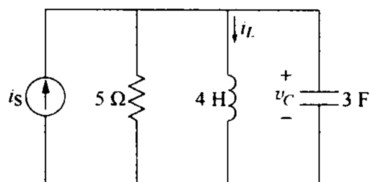
27. 如下图所示，假设开关已处于位置 a 足够长的时间。在 $t=0$ 时刻，开关被瞬时切换到位置 b，请计算 $t>0$ 时的电压 v_C 和电流 i_1。



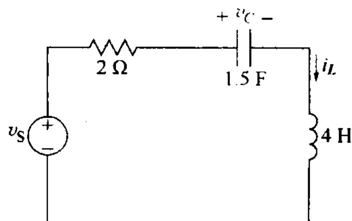
28. 如下图所示，假设开关已处于位置 a 足够长的时间。在 $t=0$ 时刻，开关被瞬时切换到位置 b，请计算 $t>0$ 时的电压 v_C 和电流 i_1。



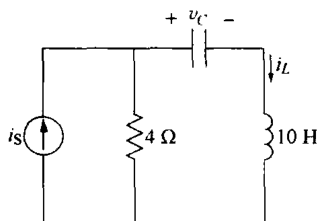
29. 假设（a）$i_S=3u(t)\text{A}$；（b）$i_S=1+3u(t)\text{A}$，请分别计算 $t>0$ 时如下电路的电流 i_L 和电压 v_C。



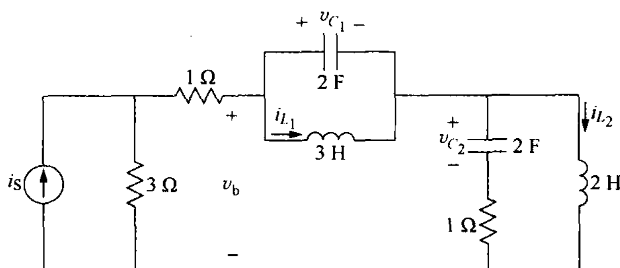
30. 假设（a）$v_S=5u(t)\text{V}$；（b）$v_S=5u(t)+3\text{V}$，请分别计算 $t>0$ 时如下电路的电流 i_L 和电压 v_C。



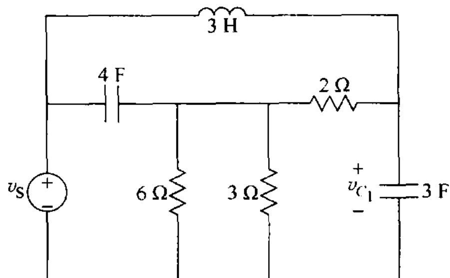
31. 假设（a）$i_S=3u(t)$A；（b）$i_S=3u(t)-1$A，请分别计算 $t>0$ 时如下电路的电流 i_L 和电压 v_C。



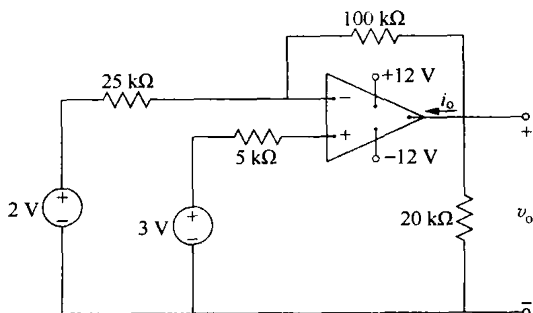
32. 假设以下电路中 $i_{L_1}(0)=2$A，$i_{L_2}(0)=5$A，$v_{C_1}(0)=2$V，$v_{C_2}(0)=-3$V，$i_S=2e^{-2t}u(t)$A。请使用节点电压法计算 $t>0$ 时的 v_b。



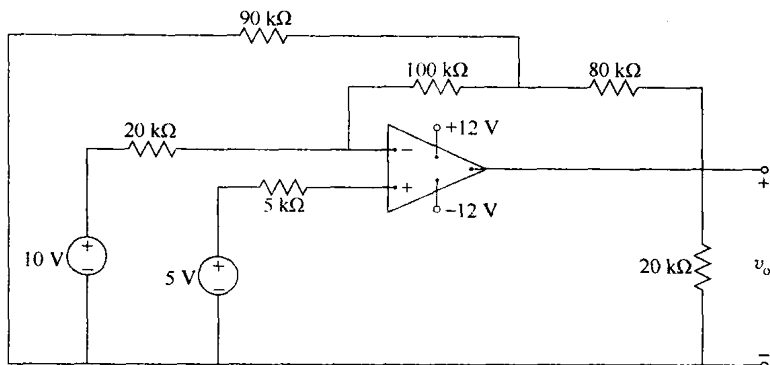
33. 假设以下电路中（a）$v_S=2e^{-3t}u(t)$V；（b）$v_S=3\cos(2t)u(t)$V；（c）$v_S=3u(t)-1$V。请分别使用节点电压法计算 $t>0$ 时的 v_{C_1}。



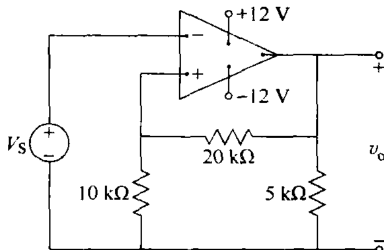
34. 假设以下电路中的运算放大器是理想的，试计算电压 v_o 和电流 i_o。



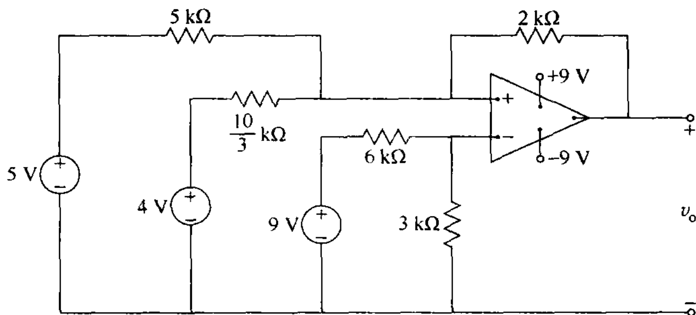
35. 假设以下电路中的运算放大器是理想的，试计算电压 v_o。



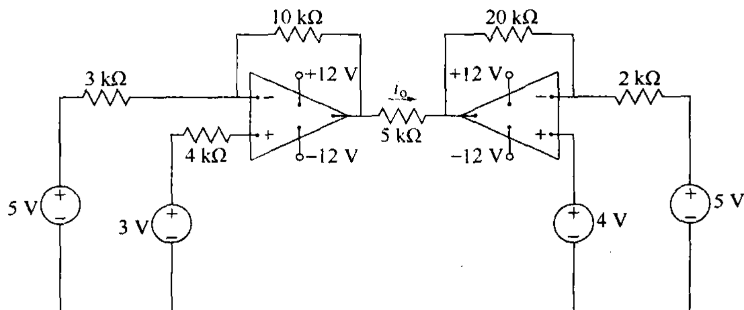
36. 如果以下电路中的运算放大器是理想的，试计算其总增益。并作出输出电压 v_o随输入电压 V_S变化的曲线，其中 V_S的取值范围是 0～10V。



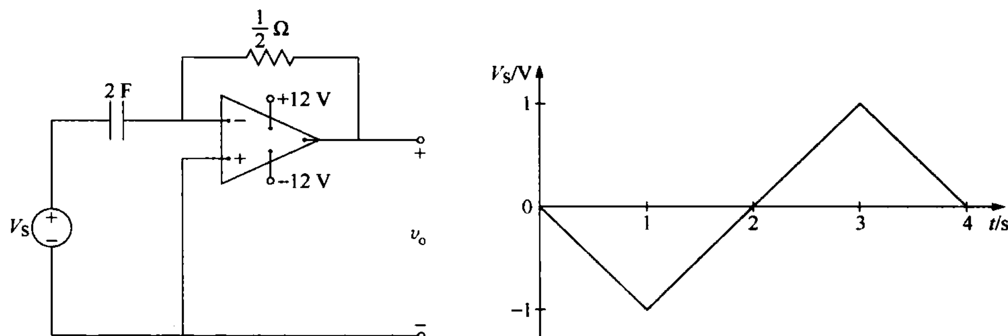
37. 假设以下电路中的运算放大器是理想的，试计算其输出电压 v_o。



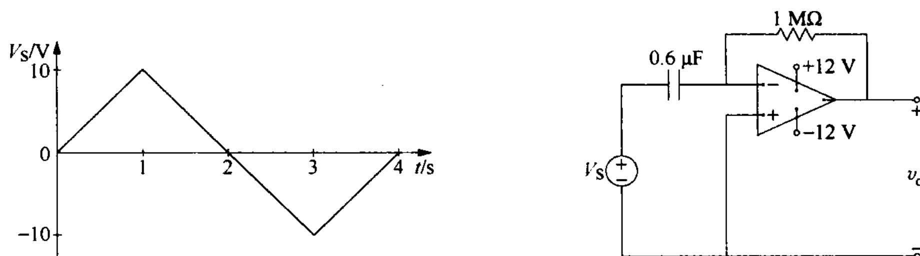
38. 假设以下电路中的运算放大器是理想的，试计算电流 i_o。



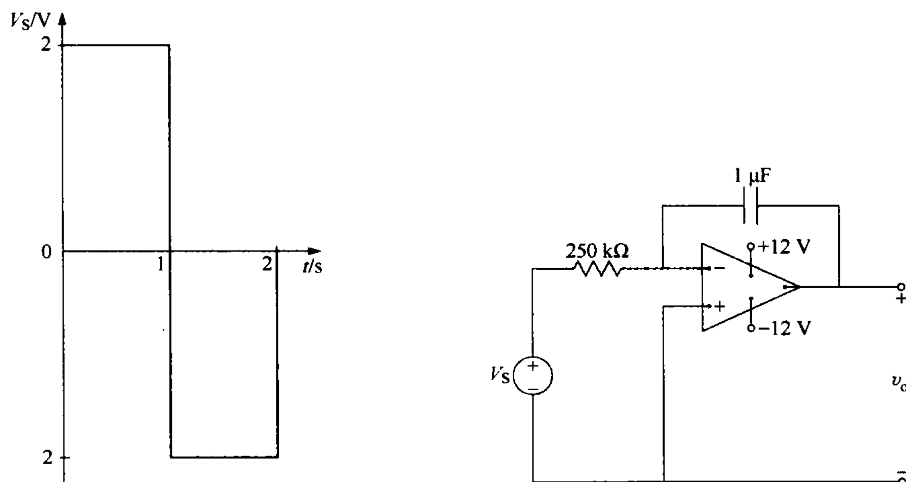
39. 设以下电路的输入 V_S为如图所示的三角波，如果电路中没有初始存储电能，并且运算放大器是理想的，请计算其输出电压 v_o。



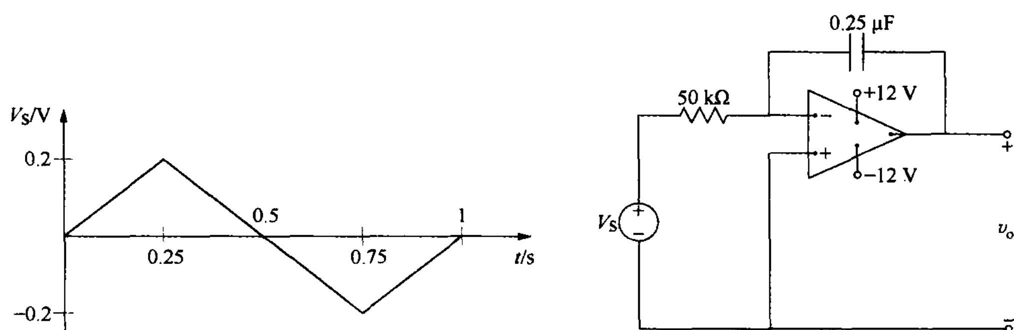
40. 设以下电路的输入电压 V_S如图所示，如果电路中没有初始存储电能，并且运算放大器是理想的，请计算输出电压 v_o。



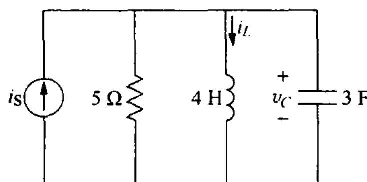
41. 设以下电路的输入电压 V_S如图所示，如果电路中没有初始存储电能，并且运算放大器是理想的，请计算输出电压 v_o。



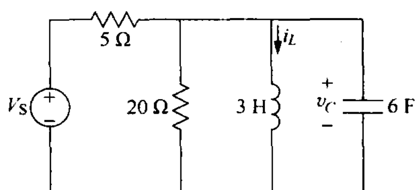
42. 设以下电路的输入电压 V_S如图所示，如果电路中没有初始存储电能，并且运算放大器是理想的，请计算输出电压 v_o。



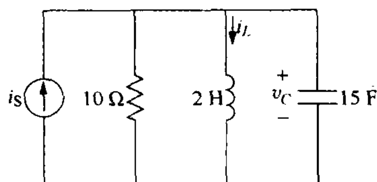
43. 设以下电路运行于正弦波稳定状态，如果 $i_S=30\cos 20t$A，试求电流 i_L 的稳态方程。



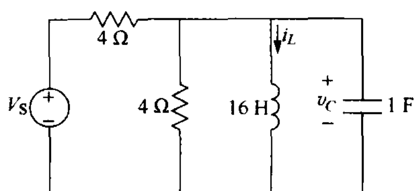
44. 假设以下电路运行于正弦波稳定状态，如果 $v_S=10\sin 1000t$V，试求电压 v_C 的稳态方程。



45. 假设以下电路运行于正弦波稳定状态，如果 $i_S=5\cos 500t$A，试求电流 i_L 的稳态方程。



46. 设以下电路运行于正弦波稳定状态，如果 $i_S=25\cos 4000t$V，试求电压 v_C 的稳态方程。



47. 试设计一个低通滤波器，使其幅值增益为 10，截止频率为 250rad/s。

48. 试设计一个高通滤波器，使其幅值增益为 20，截止频率为 300rad/s。

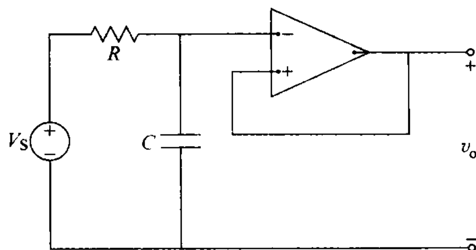
49. 试设计一个带通滤波器，使其幅值增益为 20，通频带为 50～200rad/s。

50. 试设计一个低通滤波器，使其幅值增益为 5，截止频率为 200rad/s。

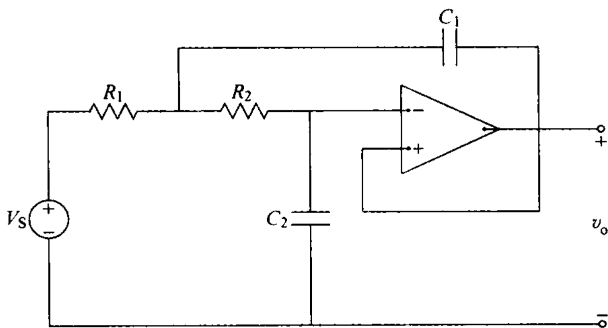
51. 试设计一个高通滤波器，使其幅值增益为 10，截止频率为 500rad/s。

52. 试设计一个带通滤波器，使其幅值增益为 10，通频带为 20～100rad/s。

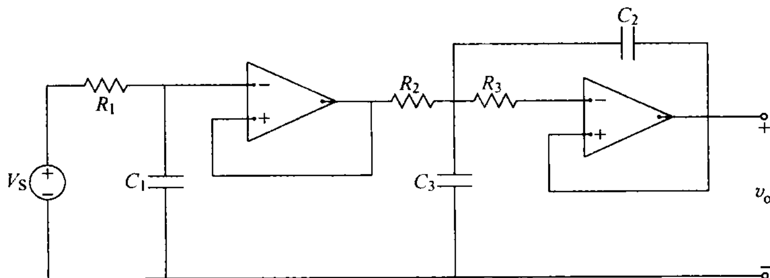
53. 假设以下电路中的运算放大器是理想的，该电路是一个一阶巴特沃思低通滤波器。试计算输出电压 v_o随频率变化的幅值特性。



54. 假设以下电路中的运算放大器是理想的，该电路是一个二阶巴特沃思低通滤波器。试计算输出电压 v_o随频率变化的幅值特性。



55. 假设以下电路中的运算放大器是理想的，该电路是一个三阶巴特沃思低通滤波器。试计算输出电压 v_o随频率变化的幅值特性。



推荐阅读资料

Aston, R. (1990). *Principles of Biomedical Instrumentation and Measurement*. Macmillan, New York.

Bronzino, J.D. (1986). *Biomedical Engineering and Instrumentation: Basic Concepts and Applications*. PWS Engineering, Boston.

Bronzino, J.D., Smith, V.H. and Wade, M.L. (1990). *Medical Technology and Society: An Interdisciplinary Perspective*. MIT Press, Cambridge, MA.

Carr J.J. and Brown, J.M. (2000). *Introduction to Biomedical Equipment Technology*, 4th Ed. Prentice Hall, Upper Saddle River, NJ.

Dempster, J. (1993). *Computer Analysis of Electrophysiological Signals*. Academic, London.

Johns, D.A. and Martin, K. (1997). *Analog Integrated Circuit Design*. John Wiley, New York.

Nilsson J.W. and Riedel, S. (2004). *Electric Circuits*, 7th Ed. Prentice Hall, Upper Saddle River, NJ.

Northrop, R.B. (2001). *Noninvasive Instrumentation and Measurement in Medical Diagnosis*. CRC, Boca Raton, FL.

Northrop, R.B. (1997). *Introduction to Instrumentation and Measurements*. CRC, Boca Raton, FL.

Perez, R. (2002). *Design of Medical Electronic Devices*. Academic, San Diego, CA.

Rosen A. and Rosen, H.D. (Eds.). (1995). *New Frontiers in Medical Device Technology*. John Wiley, New York.

Webster, J.G. (Ed.). (2003). *Bioinstrumentation*. John Wiley, New York.

Webster, J.G. (Ed.). (1997). *Medical Instrumentation—Application and Design*, 3rd Ed. John Wiley, New York.

Welkowitz, W., Deutsch, S. and Akay, M. (1992). *Biomedical Instruments—Theory and Design*, 2nd Ed. Academic, San Diego, CA.

Wise, D.E. (Ed.). (1991). *Bioinstrumentation and Biosensors*. Marcel Dekker, New York.

Wise, D.E. (Ed.). (1990). *Bioinstrumentation: Research, Developments, and Applications*. Butterworth, Stoneham, MA.

第 9 章　生物医学传感器

Yitzhak Mendelson 博士

本章目录

本章主要内容：

- 生物医学传感器的种类；
- 生物医学传感器封装材料应该具备的重要特性；
- 各种不同电极在电解液中所产生的半电池电位；
- ECG、EEG、EMG 记录电极，以及细胞内信号记录电极；
- 位移传感器、气体流量传感器以及温度传感器的基本工作原理；
- 血气和血液 pH 值的测量原理；
- 酶传感器和微生物传感器的工作原理及其应用；
- 光学传感器的工作原理及其应用。

9.1　绪论

生物医学传感器广泛用于临床医学和生物学研究中的各种生理参数测量，是各种医疗诊断仪器的重要组成部分。这类传感器可以在体侵入式或非侵入式连续监测各种重要的生理参数，也可以用于离体检测，辅助临床诊断。有些生物医学传感器还可以用于其他非医学领域，如环境监测、农业、生物加工、食品生产、石油化工和制药等领域。

在确保医疗保健服务质量的前提下，未来临床医学的发展方向是降低费用、提高效率、尽可能提供低成本高质量的服务。在这种趋势下，大量新型检测技术的不断引进，正在快速改变着临床检测的各个方面，使医生诊断和治疗疾病的方法不断变革。其中，发展最快的两个领域是病人的自我检查技术和医生的筛查技术。医生和病人都非常希望实现即时诊断，而且希望检测仪器设备能够走出医院，来到病人身边。

生物医学传感器在医学诊断中具有极其重要的作用。根据不同的需要，有些传感器主要用于临床化验室，离体检测各种生理指标，如血液中的电解质、酶等生化代谢物；有些传感器则用于在体连续监测病人的生理状态，如氧气、二氧化碳等的分压、流量和浓度等。在体实时连续监测主要用于变化速度非常快、且难以预料的生理参数。

为了准确地进行医学检测，传感器的设计和使用必须满足严格的要求。一般，新开发的生物医学传感器首先需要进行离体测试，评价其精度、测量范围、响应时间、灵敏度、分辨率和重复性⊖等指标。然后，还要根据具体的应用情况，进行在体测试，确保传感器的各项技术指标达到要求，使其能够提供灵敏、稳定、安全、高

⊖ 精度是指真实值与测量值之差除以真实值所得到的百分比；测量范围是指可以准确测量的最大值和最小值；响应时间是指达到 90%最终测量值所需的时间；灵敏度是指输出与输入的幅值变化之比；分辨率是指可以准确测得的最小幅值变化量；重复性是指重复测量同一个输入值时输出值的一致性。——原文注

效的测量。

9.1.1 传感器的分类

生物医学传感器通常按照被测参量来分类，一般可以分为物理传感器、电学传感器和化学传感器。另外，还有生物传感器，它可以看做一类特殊的生物医学传感器，具有两个独特的组成部分：酶、抗体或者受体等生物识别元件，作为中介物，它们对于特定的化学成分（即被测物）具有很强的选择性；与生物识别元件紧密结合从而构成传感器的支撑部件。这类传感器的功能就是将生化反应的结果转换为与特定化学物质浓度成正比的光、电等物理信号。根据这种分类原则，血液 pH 值传感器检测的虽然是生物参数，但它并不属于生物传感器，而只是可以测量生物参量的一种化学传感器。

9.1.2 传感器的封装

对于在体测量的生物医学传感器，封装是其设计、制作和使用过程中必须重点考虑的因素之一。当然，传感器首先必须安全可靠。但是，开发植入式传感器时，还有一个关键问题就是使用寿命和生物相容性。当传感器植入人体并与体液接触之后，人体环境就会影响传感器的工作，同时，传感器也会影响植入部位的生理功能。例如，蛋白质的吸收和细胞沉积会改变传感器封装的渗透性，而根据设计，这种渗透性既要能够保护传感器，又要允许被测化学物质在体液与传感器之间的自由扩散。一方面，如果植入式传感器的封装不合适，那么传感器的灵敏度和稳定性就会随时间的流逝发生变化，其性能会逐渐下降。另一方面，传感器植入所引起的组织炎症、感染、血管中的凝血等都会对人体造成损害（详见第 6 章）。因此，传感器外表面所用的材料对于植入式传感器的整体性能和寿命起着决定性的作用，这些材料必须无毒，并且不会造成血栓。一种比较容易的解决方法是使用各种聚合物涂料构成隔离层，将传感器和人体组织隔离，尽可能减少传感器中可能的有害成分进入人体。但是，必须记住，在常规的蒸汽、环氧乙烷和 γ 射线等消毒处理过程中，传感器封装材料的化学扩散性质绝对不能发生变化。

本章将介绍临床和科研中使用的各种生物医学传感器的工作原理，包括测量生物电和其他各种物理、化学参量的侵入式和非侵入式传感器。

9.2 生物电检测

生物电可以用各种各样的电极来检测，电极的作用就是通过耦合将生物体产生的离子电位传给电子仪器。电极可以分为非侵入式的体表电极和侵入式的微电极两类。

9.2.1 电解质与金属电极之间的界面

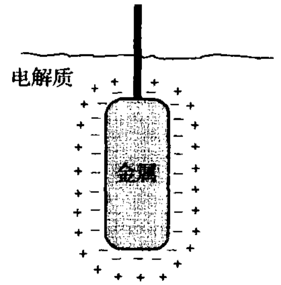


图 9.1　金属与电解质界面上的电荷分布

如果将金属放入电解质溶液，其表面就会形成如图 9.1 所示的电荷分布界面。这种局部电荷分布会产生一个电压差，该电压差就是建立在金属与电解质溶液之间的半电池电位。

表 9.1 列出了几种重要金属材料的半电池电位，注意，氢电极是一种标准电极，其他金属的半电池电位都以其为标准测得。

表 9.1　几种重要金属材料的半电池电位

金属及其化学反应	半电池电位
$Al \to Al^{3+} + 3e^-$	−1.706
$Cr \to Cr^{3+} + 3e^-$	−0.744
$Cd \to Cd^{2+} + 2e^-$	−0.401
$Zn \to Zn^{2+} + 2e^-$	−0.763
$Fe \to Fe^{2+} + 2e^-$	−0.409
$Ni \to Ni^{2+} + 2e^-$	−0.230
$Pb \to Pb^{2+} + 2e^-$	−0.126
$H_2 \to 2H^+ + 2e^-$	0.000（定为标准）
$Ag \to Ag^+ + e^-$	+0.799
$Au \to Au^{3+} + 3e^-$	+1.420
$Cu \to Cu^{2+} + 2e^-$	+0.340
$Ag + Cl^- \to AgCl + 2e^-$	+0.223

例 9.1　将银电极和锌电极同时浸入电解质溶液中，请计算两个电极之间的电压差。

解：由表 9.1 可知，银电极和锌电极的半电池电位分别为 0.799V 和 −0.763V。因此，这两种金属电极之间的电压差为

$$0.799V-(-0.763V)=1.562V$$

一般，生物电的测量都采用同一种金属制作的相同电极，这样，两个电极产生的半电池电位的大小就相等。例如，在胸前贴上两个相同的电极就可以测量心脏所产生的电压信号，也就是心电图（ECG）。理想情况下，假设两个电极上产生的皮肤与电极界面处的电位相等，那么连接在两个电极上的差动放大器就只放大心电信号，而半电池电位则被相互抵消。但是，实际测量时，由于电极材料的差别和皮肤接触阻抗的不同，会造成一个明显的直流偏置电压，并在两个电极之间形成电流，该电流流过人体时会产生电压降，叠加在放大器的输出信号上，造成记录信号的不稳定和基线漂移。■

例 9.2　将银电极和铝电极同时浸入电解质，如果电解质溶液的等效阻抗为 2kΩ，请计算流过两个电极的电流。

解：计算如下：

$$0.799\text{V}-(-1.706\text{V})=2.505\text{V}$$
$$2.505\text{V}/2\text{k}\Omega=1.252\text{mA}$$

■

9.2.2 心电图记录电极

图9.2所示是两种非侵入式ECG记录电极。其中一种柔软的ECG电极用弹性橡胶聚合物制作，聚合物中加入了极细的碳粉和金属粉末，使其具有较好的导电性（见图9.2a）。使用时，加上AgCl导电膏，用双面胶将电极固定在皮肤上就可以测量，既快速又方便。最常用的生物电记录电极是银/氯化银电极（即Ag/AgCl），它是在银电极表面通过电化学反应镀上一层很薄的氯化银薄膜制作而成（见图9.2b）。为了使电极与皮肤之间形成良好的电接触，这种电极通常嵌在吸盘内，并浸没在吸有电解质导电液的塑料泡沫中。饱浸电解质溶液的泡沫材料还可以减少皮肤与电极表面相对移动引起的伪迹，这种移动伪迹可以造成很大的干扰信号，甚至会破坏检测信号。

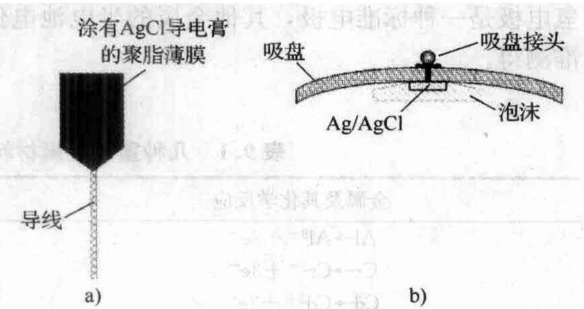


图9.2 体表ECG电极
a）柔软的聚酯薄膜电极 b）一次性吸盘式Ag/AgCl电极

9.2.3 肌电图记录电极

记录人体不同肌肉组织肌电图（EMG）的电极有很多种，电极的电特性和适用的记录部位各不相同，所记录的EMG信号波形和幅值也不同。对于非侵入式记录，皮肤表面的预处理很重要，通常要用酒精清洁皮肤，或者使用少量电解质导电膏，有助于减小皮肤与电极表面之间的接触阻抗，改善记录信号的质量。通过体表EMG记录测试神经传导功能时，最常用的电极呈圆盘状，直径约1cm，由银或铂制作。如图9.3所示，经皮穿刺的电极针用于侵入式的神经纤维和肌肉纤维电信号的直接测量。最常见的电极针是如图9.3a所示的双极同心电极，它由装在套管或注射器针头内的两根细金属丝构成，这两根金属丝分别作为记录电极和参考电极。还有一种经皮穿刺的EMG电极是如图9.3b所示的单极电极针。这种电极只有一根外包聚四氟乙烯塑料绝缘薄层的细金属丝，针尖暴露0.3mm左右。与双极电极不同，

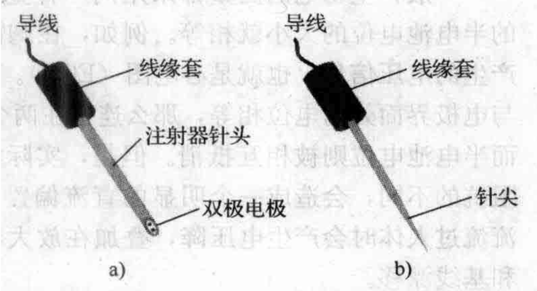


图9.3 侵入式肌电图记录电极
a）双极电极 b）单极电极

单极电极使用时还需要另外加一根参考电极，从而在两根电极之间构成闭合回路。这根参考电极要么放在记录电极附近，要么贴在皮肤表面上。

9.2.4　脑电图记录电极

最常用的脑电图（EEG）记录电极是杯状电极和皮下针状电极。杯状电极由铂或锡制成，直径 5～10mm。电极杯内充满电解质导电膏，使用时用胶带固定在头皮表面即可。不过，由于头发和皮肤油脂容易造成电极与皮肤之间的接触不良，头皮外记录 EEG 还是比较困难，因此，医生有时不用皮肤表面的金属电极，而是喜欢用皮下针状 EEG 电极。这种电极用细铂丝或者不锈钢针制作，长约 10mm，直径约 0.5mm，可以插入皮肤，获得较好的电接触。

9.2.5　微电极

微电极是具有超细锥形尖端的生物电记录电极，可以插入细胞内，记录单细胞动作电位，常用于神经生理学研究。为了尽可能减小细胞受到的损伤，电极尖端必须比细胞小很多，另一方面其尖端还必须具有足够的强度，以便能够刺穿细胞膜。图 9.4 显示了 3 种常见微电极的结构：玻璃微电极、金属微电极、固态微电极阵列。

图 9.4a 所示玻璃微电极的制作过程是：将一根玻璃毛细管（通常直径为 1mm）放在微型电加热器上加热，将其中段软化，然后在两端加力，快速将毛细管拉断，这样就可以得到两根同样的电极，电极尖端开口处的直径仅为 0.1～10μm。再在毛细管内注入浓度为3mol/L 的 KCl 溶液，并插入一小段 Ag/AgCl 导线，与电解质溶液形成电接触。当微电极尖端插入细胞内电解质中时，离子电流就可以流过电极尖端处的液接界面，从细胞内到达微电极的 Ag/AgCl 细丝，从而构成闭合电路。

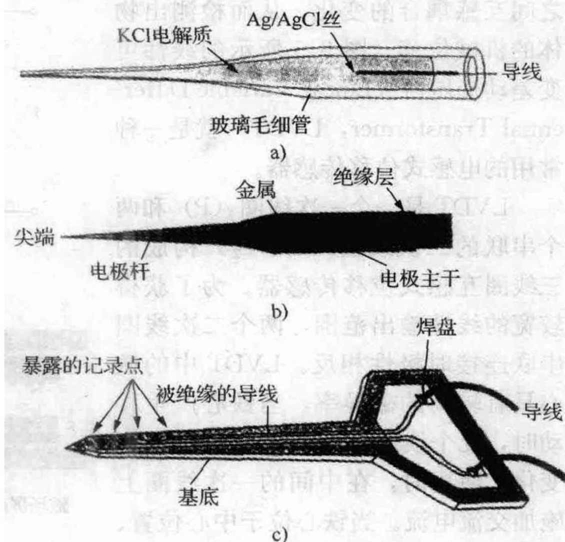


图 9.4　微电极

a）毛细玻璃管微电极　b）绝缘金属丝微电极　c）固态微电极阵列

图 9.4b 所示是另一种微电极，由极细的钨丝或者不锈钢丝等硬金属丝制作，电极的尖端通过电化学腐蚀技术做得非常尖锐，直径只有几微米。尖端之外的金属丝用绝缘层包裹。

利用集成电路的固态微加工技术还可以制作多通道微电极阵列，用于大脑和脊髓神经元的电信号记录或者电刺激。图 9.4c 所示就是这种电极的一个例子，电极包

括一块精细微加工制成的硅基底薄片，上面有 4 个暴露的记录点。这种微加工制作技术的主要优点是能够批量生产重复性很好的高质量精密微电极。

9.3 物理量的测量

9.3.1 位移传感器

电感式位移传感器

电感式位移传感器的基本原理基于线圈的电感 L，其计算公式如下：

$$L = n^2 \times G \times \mu \tag{9.1}$$

式中　G——几何形状系数；

n——线圈匝数；

μ——线圈内部磁心材料的磁导率。

这种传感器通过装在线圈内的铁氧体铁心的移动，引起单个线圈的自感变化，或者引起两个以上固定线圈之间互感耦合的变化，从而检测出物体的机械位移。图 9.5 所示的线性可变差动变压器（Linear Variable Differential Transformer，LVDT）就是一种常用的电感式位移传感器。

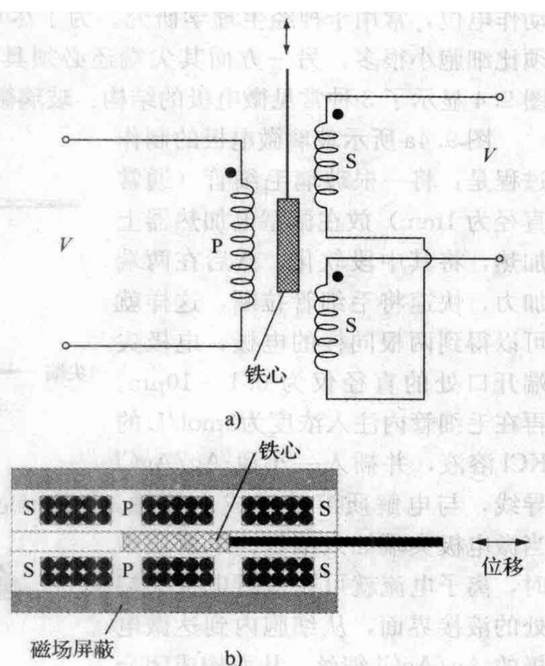


图 9.5　线性可变差动变压器（LVDT）位移传感器

a）电路示意图　b）截面图

LVDT 是一个一次线圈（P）和两个串联的二次线圈（S_1、S_2）构成的三线圈互感式位移传感器。为了获得较宽的线性输出范围，两个二次线圈串联连接时极性相反。LVDT 中的铁心具有较高的磁导率，当铁心产生移动时，几个线圈之间的互感耦合就会变化。使用时，在中间的一次线圈上施加交流电流。当铁心位于中心位置、处于两侧的两个二次线圈中的部分对称相等时，一次线圈在两个二次线圈中感应产生的电压大小相等，方向相反，两者相互抵消，净输出电压为 0。当铁心移向某个二次线圈时，该线圈的感应电压就随铁心的位移量成比例增大，而另一个二次线圈中的感应电压则按比例减小，从而形成如图 9.6 所示的电压—位移关系曲线。由于两个二次线圈的感应电压相位相反，所以必须采用专用的相位敏感放大器才能同时检测出铁心位移的大小和方向。

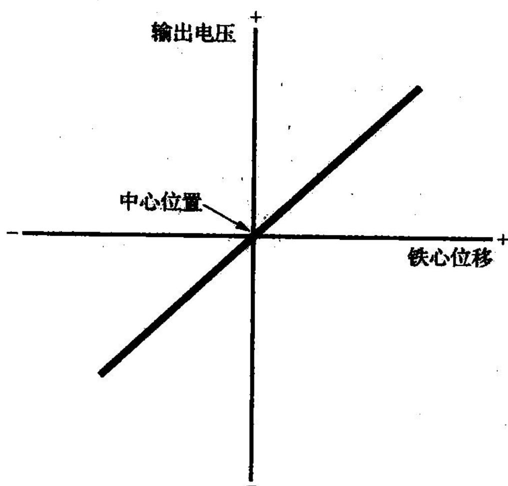


图 9.6 常用 LVDT 传感器输出电压与铁心位移之间的关系

有一种电磁感应式流量传感器，在心脏手术过程中，可用于测量暴露的主动脉等大血管中的血流速度。

假设血管直径为 l，血流速度均匀，且为 $\vec{u}$。如果将血管放置在磁场方向与血流方向垂直的均匀磁场 $\vec{B}$ 中，那么，血液中带正电的阳离子和带负电的阴离子就会受到与磁场方向和血流方向都垂直的作用力 $\vec{F}$，其计算公式为

$$\vec{F}=q(\vec{u}\times\vec{B}) \tag{9.2}$$

式中 q——单位电荷量，即 1.6×10^{-19}C。

这样，血液中的两种带电粒子就会按照作用力 $\vec{F}$ 的方向，沿着血管的径向分别向两个相反方向运动。这种运动会产生一个反作用力 $\vec{F}_O$，即

$$\vec{F}_O=q\times\vec{E}=q\times\frac{V}{l} \tag{9.3}$$

式中 $\vec{E}$——带电粒子移动时产生的净电场；
V——血管上产生的电压差；
l——血管直径。

平衡时，以上两个作用力相等，因此，电压差 V 正比于血管中血液的速度，它等于

$$V=B\times l\times u \tag{9.4}$$

例 9.3 如果用电磁流量计的探头测量血管中血流的速度，请计算探头中感应的电压差。假设血管直径为 0.5cm，血流速度为 5cm/s，磁场强度 B 为 1.5×10^{-5} Wb/m^2。

解： 由式（9.4）可得

$$V=B\times l\times u=(1.5\times10^{-5}\,\text{Wb/m}^2)\times(0.5\times10^{-2}\,\text{m})\times(5\times10^{-2}\,\text{m/s})$$
$$=3.75\times10^{-9}\,\text{V}$$

注意：[Wb]=[V×s] ■

如图 9.7 所示，电磁流量计配有夹持式探头，可以紧贴血管外周。探头中的线圈通有交流电，所产生的磁场与血流方向垂直。探头上装有一对很小的电极，紧贴血管外壁，用于记录血管上感应的电压，该电压与线圈的交流励磁电压具有相同的频率。这种

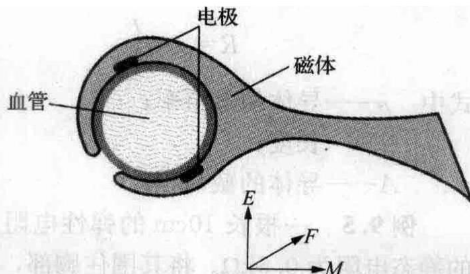


图 9.7 电磁感应式流量计

流量计不用直流励磁电压，而用交流励磁电压，是因为交流电有助于消除电极与血管壁之间接触阻抗引起的偏置电压误差。

电阻式位移传感器

电位器是一种电阻式位移传感器，通过移动电阻元件表面的滑动触点，可以将线性位移或者角位移转换成输出电压信号。图 9.8 显示了这种电位器传感器的原理。在电阻 R 上施加一个电压 V_i，滑动触点与电阻某个末端之间的电压 V_o与该两点之间的距离成正比。一般，该可变电阻上流过的电流恒定不变，其输出电压的微小变化可以根据欧姆定律，用高灵敏度电压表测得。

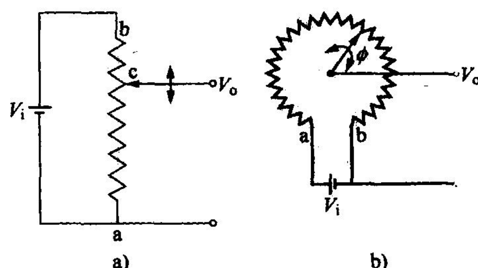


图 9.8 线性位移（图 a）和角位移的电位器传感器（图 b）

例 9.4 假设位移的变化量为 20%，请计算线性电位器传感器的输出电压变化量。

解：假设传感器中的电流恒定不变，由欧姆定律可得：

$$\Delta V = I \times \Delta R$$

由此可见，由于滑动触点与电阻某个末端之间的电阻与距离成正比，因此，20%的位移变化也将产生 20%的输出电压变化。 ■

临床上，有时需要用血压袖袋暂时阻断某条腿的静脉血流回流，来测定这条腿的外周血容量的变化。这种血容量测量方法称为体积描记法（plethysmography），它能够检查腿部静脉中是否存在较大的血栓。测量时用弹性电阻式传感器裹住大腿，测量传感器电阻随时间的变化率，该变化率对应于腿部血容量的相对变化。如果血管中存在血凝块，那么解除袖袋的阻断之后，存积在腿部的血液通过静脉回流所需的时间就会加长。同理，如果用弹性电阻式传感器软带裹住胸部，就可以检测病人的呼吸模式。

如图 9.9 所示，弹性电阻式传感器由一条充满导电材料的细弹性软管构成，软管导体的电阻可用下式计算：

$$R = \rho \times \frac{l}{A} \tag{9.5}$$

式中 ρ——导体的电阻率；

l——长度；

A——导体的截面积。

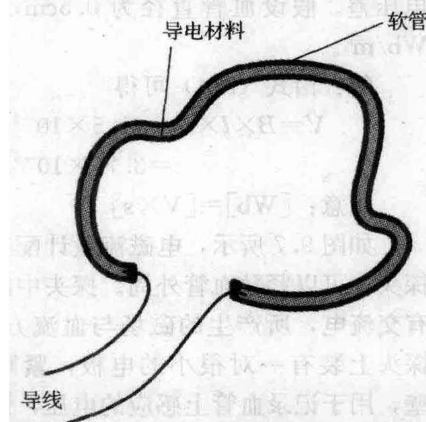


图 9.9 弹性电阻式传感器

例 9.5 一根长 10cm 的弹性电阻式传感器的静态电阻为 0.5kΩ。将其围住胸部，假设呼气时胸部直径为 33cm，请计算传感器围住胸部之

后的电阻。

解：将弹性传感器拉长，绕胸部一周后，其长度从原来的 10cm 增加到 103.7cm。假设传感器拉长之后其截面积仍然保持不变，那么，其电阻值将增加到

$$R_{(拉长后)}=0.5\text{k}\Omega\times\left(\frac{103.7\text{cm}}{10\text{cm}}\right)=5.18\text{k}\Omega$$ ■

例 9.6　假设传感器上施加的直流励磁电流为 5mA，并且正常呼吸时胸腔周长的变化为 10%。请计算例 9.5 中弹性传感器上所产生的电压变化。

解：根据欧姆定律（$V=I\times R$）：

$$V=5\text{mA}\times 5.18\text{k}\Omega=25.9\text{V}$$

如果 R 增加了 10%，那么

$$V=5\text{mA}\times 1.1\times 5.18\text{k}\Omega=28.5\text{V}$$

$$\Delta V=2.6\text{V}$$ ■

应变式位移传感器

应变片是一种位移传感器，可用于测量物体受力后的长度变化。这种传感器所产生的电阻变化与物体长度的变化率成正比。长度变化称为应变 S，定义为

$$S=\frac{\Delta l}{l} \tag{9.6}$$

式中　Δl——物体长度的微小变化量；

l——物体的原始长度。

电阻丝和某些半导体材料都可以用做应变片。

下面解释应变片的工作原理。假设有一根细导线，长为 l，截面积为 A，电阻率为 ρ；那么，根据式（9.5）可以求得其应变之前的初始电阻。设在弹性限度范围内将其拉长 Δl，则其长度变为（$l+\Delta l$）。由于拉长后导线的体积仍然保持不变，因此，长度的增加必然导致截面积的减小，即

$$lA=(l+\Delta l)\times A_{(拉长后)} \tag{9.7}$$

则拉长后导线的电阻为

$$R_{(拉长后)}=\rho\times\frac{l+\Delta l}{A_{(拉长后)}} \tag{9.8}$$

电阻的增量为

$$\Delta R=R_{(拉长后)}-\rho\times\frac{l}{A} \tag{9.9}$$

将式（9.7）和式（9.8）代入式（9.9），可得

$$\Delta R=\rho\times\frac{(l+\Delta l)^2}{l\times A}-\rho\times\frac{l}{A}=\frac{\rho\times(l^2+2l\Delta l+\Delta l^2-l^2)}{l\times A} \tag{9.10}$$

假设长度变化量很小，即 $\Delta l\ll l$，则上式可简化为

$$\Delta R=\rho\times\frac{2\times\Delta l}{A}=\frac{2\times\Delta l}{l}\times R \tag{9.11}$$

电阻变化率（$\Delta R/R$）与长度变化率（$\Delta l/l$）之比称为应变系数 G。一般金属丝

应变片的应变系数是常数，其值约为 2；而硅材料制作的半导体应变片的灵敏度要高得多，其应变系数是金属丝的 70～100 倍。

例 9.7 请计算电阻变化 10%时，电阻丝应变片的应变大小。

解：由式（9.6）和式（9.11）可得

$$\frac{\Delta R}{R}=\frac{2\times\Delta l}{l}=2\times S$$

即

$$0.1=2\times S$$
$$S=0.05$$

■

应变片通常可分为粘合式和非粘合式两种。如图 9.10 所示，粘合式应变片由一根盘绕的细导线粘贴在半弹性基片上构成。非粘合式应变片由多根电阻丝构成（一般是 4 根），这些电阻丝跨接在固定支架与活动支架之间。当应变片受到外力作用时，4 根电阻丝中总是有两根被拉伸，而另两根则被压缩。如图 9.11 所示，用这种应变片可以制作血压传感器，传感器中有一片弹性膜片通过枢轴直接连接在活动支架上。外周血管中的血液通过一根灌满生理盐水的细导管与一次性弹性帽相连，弹性帽的底部就是弹性膜片封口。心脏搏动的血压变化对膜片产生作用力，使得与膜片相连的活动支架偏离静息位置，这个位移引起应变片的伸缩，从而产生与血压搏动成比例的周期性电阻变化。

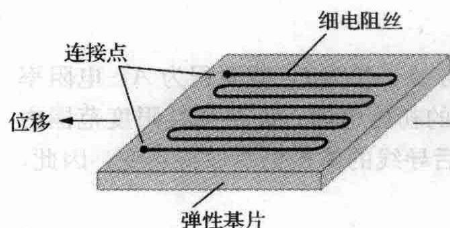


图 9.10 粘合式应变片传感器

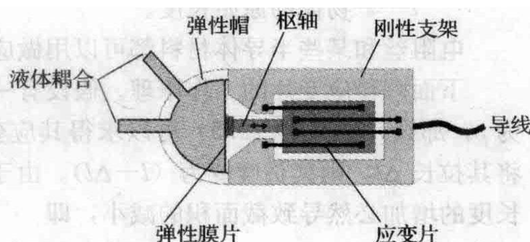


图 9.11 电阻应变式（非粘合式）血压传感器

应变片的电阻变化一般非常小，另外，温度变化也会造成电阻丝的热膨胀，引起应变片的电阻变化。因此，应变式传感器通常需要配置具有温度补偿电路的高灵敏度放大器。

电容式位移传感器

两块大小相同的平行极板构成的电容，其电容值 C 等于

$$C=\varepsilon_0\times\varepsilon_r\times\frac{A}{d} \tag{9.12}$$

式中 A——平行极板的面积；

d——两块平行极板的间距；

ε_0——真空介电常数，$\varepsilon_0=8.85\times10^{-12}\text{F/m}$；

ε_r——极板间绝缘介质的相对介电常数。

如图 9.12a 所示，电容式传感器测量位移的最常用方法是检测固定极板与活动极板之间的距离变化。这种传感器可用于测量力、压力和加速度。另外，如图 9.12b 所示，还可以在这种结构中再加入第三块极板，构成差动电容式传感器。其外侧的两块极板是固定的，中间一块极板则是活动的，它可以在两块固定极板之间自由移动，从而构成了两个可变电容器。显然，中间极板的位移将使其与两块固定极板的间距分别变化 $+\Delta d$ 和 $-\Delta d$，从而使一个电容的值增加，而使另一个电容的值减小。这种双电容结构可以提高传感器的灵敏度，并且可以组成惠斯登电桥。利用制造集成电路的固态微加工技术很容易批量生产电容式传感器。

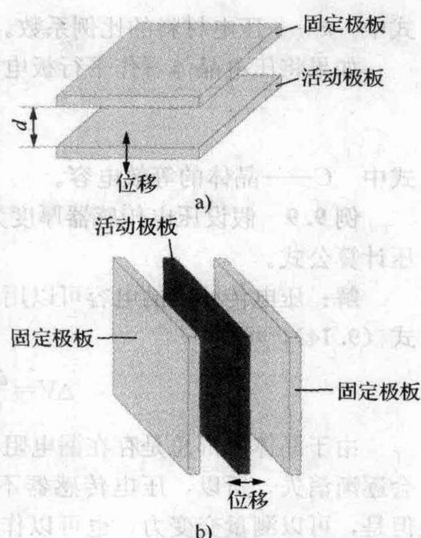


图 9.12　电容式位移传感器
a) 单电容　b) 差动式双电容

例 9.8　假设某电容式传感器的两块金属极板的面积均为 4cm²，间距为 0.1mm，板间介质的介电常数为 $\varepsilon_r=2.5$，请计算该传感器的电容值。

解：$C=\varepsilon_0\times\varepsilon_r\dfrac{A}{d}=8.85\times10^{-12}\text{F/m}\times2.5\times4\times10^{-4}\text{m}^2/(0.1\times10^{-3}\text{m})=0.885\text{F}$ ■

如果将多个电容式位移传感器安装在床垫里，就可以测量人体的呼吸运动。如果将传感器的活动极板与接触液体或气体的薄膜相连，还可以制作成压力传感器，这时只要在电容上加上电压，并用放大器放大薄膜移动时所产生的微小信号，就可以获得与被测压力成正比的输出电压信号。

压电式位移传感器

压电传感器常用于心音和血压等各种力和加速度的生理参数测量，也用于产生超声波（即频率超过 20kHz 的高频声波），超声波可用于血流测量以及人体软组织成像（详见第 16 章）。

压电传感器中有一块很小的石英晶体，如图 9.13 所示，如果在晶体的两块极板上外加一个电场，通常是一个窄脉冲电压，那么，晶体就会收缩。反过来，如果通过机械作用使晶体产生应变，则其两极板之间会产生微弱的电压信号。除了石英晶体以外，钛酸钡和锆钛酸铅等许多陶瓷材料也会产生这种压电效应。

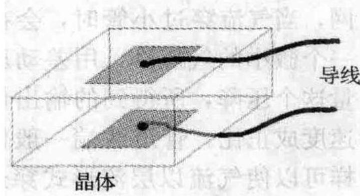


图 9.13　压电晶体传感器

压电效应的产生机理是：晶体在外力作用下，其非对称晶格产生变形，内部正、负电荷方向发生变化，重新排布，从而使晶体两个端面上产生表面电荷 Q，此电荷的量与受力大小成正比，即

$$Q=k\times F \tag{9.13}$$

式中　k——压电材料的比例系数。

如果将压电晶体看作平行板电容器，那么，晶体两端的电压 ΔV 等于

$$\Delta V=\frac{\Delta Q}{C} \tag{9.14}$$

式中　C——晶体的等效电容。

例 9.9　假设压电传感器厚度为 d，面积为 A，所受外力为 F，请推导其输出电压计算公式。

解：压电传感器的电容可以用式（9.12）计算。将式（9.12）和式（9.13）代入式（9.14），可得

$$\Delta V=\frac{\Delta Q}{C}=\frac{k\times F}{C}=\frac{k\times F\times d}{\varepsilon_0\times\varepsilon_r\times A}$$ ■

由于晶体内部总是存在漏电阻，因此，其两个端面上产生的稳态电荷最终必然会逐渐消失，所以，压电传感器不能用于测量恒定的或者频率很低的直流型外力，但是，可以测量交变力，也可以作为机械共振器，发出 1～10MHz 的高频信号，用做晶体振荡器和超声脉冲发生器。

在生物医学工程中，压电传感器常用于测量物体的厚度，或者非侵入式监测血压等。例如，将两块同样的晶体放在血管两侧，并在其中一块晶体上施加电源，使其产生一小段超声波脉冲信号，超声波脉冲传播到另一块晶体上并被接收，这样，通过测量超声波在两个传感器之间的传播时间，就可以计算出血管管径。假设超声在软组织中传播的速度 c_t 已知（一般为 1500m/s），并且测得的超声波传播时间为 t，那么，就可以利用下式计算出两个传感器之间的距离：

$$d=c_t\times t \tag{9.15}$$

9.3.2　气体流量传感器

最常见的医用气体流量传感器是 Fleish 呼吸气体流量计（Fleish pneumotachometer）。如图 9.14 所示，它由一段短直管构成，管道中间部位装有一个固定的筛状隔网，当气流穿过小管时，会在隔网上产生一个微小的气压降，用差动压力传感器测量这个压降，所得到的输出信号就与气流速度成正比。管道两端一般做成锥形，这样可以使气流以层流形式穿过管道。为了避免在长时间使用过程中，水汽在隔网上冷凝聚集，从而干扰气压降的测量，需要用一个小型加热器加热隔网。病人使用机械呼吸机通气时，常用这种 Fleish 呼吸流量计监测病人的呼吸容量、流量和呼吸频率。

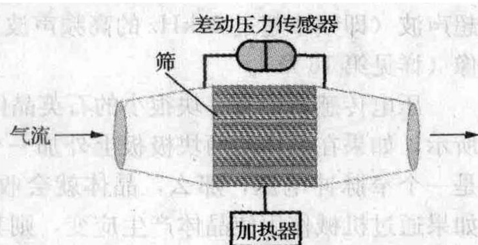


图 9.14　Fleish 呼吸气体流量计

9.3.3 温度测量

体温是人体调控最严密的生理参数之一，也是日常生活中衡量病人健康状况的四大基本指标之一。人体深部温度非常稳定，健康人大约为 37℃，变化范围一般在 ±0.5℃之内。因此，体温升高往往就意味着患病或者感染，而皮肤温度明显下降可能是临床休克的征兆。

人体体温的测量通常在体表或者体腔内进行，如腋下体表测温，口腔或者肛门等体腔内测温。常用的体温计有接触式和非接触式两种，热敏电阻体温计必须与人体皮肤或粘膜组织直接接触才能测量体温，而耳道体温计就是一种非接触式测量人体深部温度的体温计。

热敏电阻用压缩烧结的镍、锰、钴等金属氧化物制成，其电阻值会随温度的变化而变化。如图 9.15 所示，市场上出售的热敏电阻有较小的珠形到较大的盘形等多种形状。



图 9.15　常见热敏电阻的形状

热敏电阻的电阻－温度特性曲线可以用以下数学公式表示：

$$R_T = R_0 \times \exp\left[\beta \times \left(\frac{1}{T} - \frac{1}{T_0}\right)\right] \tag{9.16}$$

式中　R_0——在参考温度 T_0（是热力学温度，单位为 K）时的电阻值；

R_T——在温度 T（是热力学温度，单位为 K）时的电阻值；

β——热敏电阻系数，一般取值为 2500～5500K。

图 9.16 显示了热敏电阻典型的电阻一温度特性曲线。注意，热敏电阻具有非线性的温度—电阻关系，并且其温度系数为负，也就是温度升高时，其阻值反而减小。而一般的金属和常规电阻的温度系数则为正，也就是温度升高时，元件的电阻值也增大。

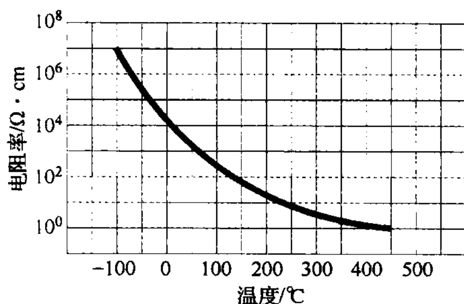


图 9.16　热敏电阻的典型电阻一温度特性曲线

例 9.10　某热敏电阻温度计的热敏系数 β 为 4500K，假设其在 37℃ 体温下的电阻值为 85Ω，请计算 25℃ 时的电阻值。

解：利用式（9.16）的热敏电阻的电阻一温度特性公式，可得

$$R_T = 85\Omega \times \exp\left[4500 \times \left(\frac{1}{298} - \frac{1}{310}\right)\right] = 152.5\Omega$$ ■

医用温度计的热敏电阻探头必须做得很小，这样它对温度变化的响应速度就比较快。使用时，探头要用一次性弹性杀菌薄膜包住，以防止病人之间的交叉感染。

热敏电阻传感器可用于Swan-Ganz导管热稀释法测定心输出量（即每分钟从心脏泵出的血容量），从而评价心室功能。这种测量通常在手术室或重症监护室进行。如图9.17b所示，测量时需要通过软导管向右心房快速注入低温示踪剂冷溶液，一般是3～5mL经过灭菌的零度生理盐水或者葡萄糖溶液。

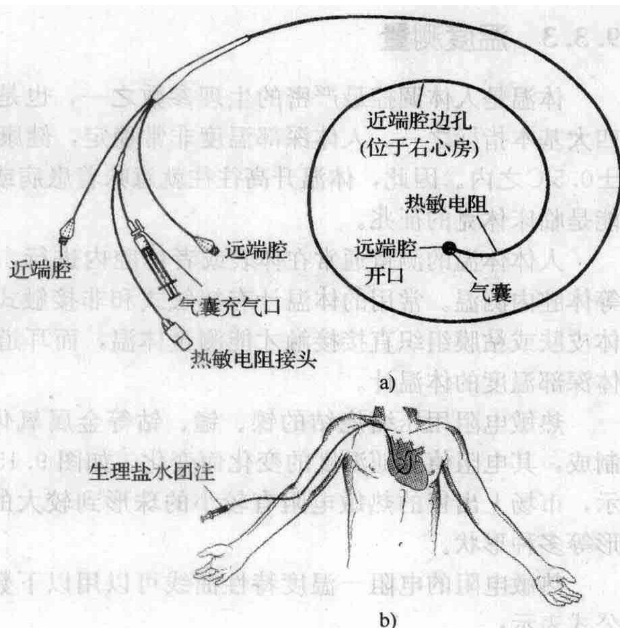


图9.17 Swan-Ganz导管

如图9.17a所示，Swan-Ganz导管有5号或7号两种，靠近导管头部处有一个小气囊，可以引导导管经股静脉或颈内静脉进入心脏。导管一般用X射线不能透过的材料制成，这样，就可以在X射线影像上方便地观察导管的位置。常用的Swan-Ganz导管是三腔管，其中，气囊腔用于引导导管的头部进入正确的位置，另外还有开口在近端的中心静脉腔和开口在远端的肺动脉腔。充胀后的气囊可以引导导管头部穿过三尖瓣进入右心室，然后再穿过肺动脉瓣，进入肺动脉。如果将导管的近端腔和远端腔两根管子的出口连接到压力传感器上，那么，当导管进入正确的位置之后，就可以监测肺动脉和右心房的血压。

导管插入后，导管近端腔的开口刚好位于右心房，因此，通过近端腔可以向右心房团注冷溶液。注射的溶液随即混入右心房血液中，使血液温度有所下降。然后，这些降温的血液被右心室射入肺动脉，位于肺动脉的导管远端腔的管壁上装有热敏电阻，可以监测血流的温度变化。仪器会自动计算血液温度变化曲线的积分值，并由此求出心输出量，血液温度变化的积分值与心输出量成反比。

非接触式耳道体温计可以检测耳朵鼓膜附近的温度，该处的温度与人体深部的温度之间保持0.5～1.0℃的恒定温差。如图9.18所示，其原理是：鼓膜发出的红外辐射通过具有高反射率镀金内表面的金属导波管，传导给热敏探测器。热敏探测器可以是热电堆，也可以是热电传感器，它将热转化为电。探测器一般置于恒温环境中，以减少环境温度波动引起的误

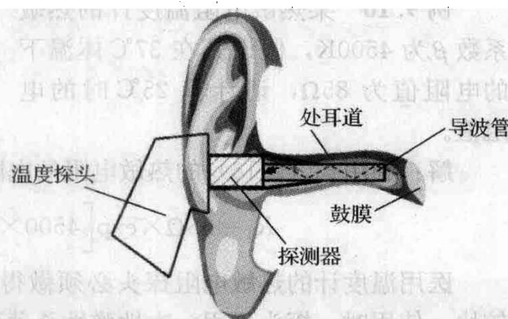


图9.18 非接触式红外耳道体温计

差。体温计探头上使用一次性耳镜，以防止交叉感染。

9.4　血气传感器和 pH 计

手术室和重症监护室中经常需要监测危重病人动脉血液的氧气和二氧化碳分压（即 P_{O_2} 和 P_{CO_2}）以及 pH 值，这些指标可以反映人体呼吸系统问题和代谢失调，并且反映血液的氧气补充和二氧化碳排出是否正常。医生根据这些指标来调节机械呼吸机的设置，或者调整用药。

常规血气分析一般都要从外周动脉抽取血样，然后将血样送到医院化验室进行分析。为了实现病情不稳定的危重病人的快速检测，人们开发了各种非侵入式连续监测血气指标的新方法，使医生可以随时跟踪病人的状况，并能够及时获得治疗效果的反馈信息。

非侵入式动脉血液氧气和二氧化碳检测的依据是，这两种气体都能够较容易地穿过皮肤扩散到体外。由于皮肤下浅表血管中血液的气体分压与皮肤表面外的气体分压之间存在压差，因此，就会产生扩散。根据这一原理，人们已经研制出两种非侵入式经皮监测 P_{O_2} 和 P_{CO_2} 的电化学传感器。另外，血液呈现的颜色与红细胞中血红蛋白的含氧量相关，根据这一现象，人们还开发了多种测量血氧饱和度的光学方法。

9.4.1　氧气的测量

血氧的定量监测对于评价病人血液循环系统和呼吸系统的状况非常重要。通过血液循环系统，肺部的氧气被运输到全身各个组织。在血液中，氧气有两种不同的存在形式。正常生理状态下，血液携带的总氧气量中大约只有 2%溶解于血浆，这个比值与血氧分压 P_{O_2} 成正比。其余 98%氧气则存在于红细胞中，并通过可逆化学键的松散连接，与还原血红蛋白 Hb 结合成为氧合血红蛋白 HbO_2。因此，可以用两种方法测量血液的氧含量：一种是氧分压极谱分析法，另一种是血氧饱和度的光电检测。血氧饱和度 S_{O_2} 是指血液中 HbO_2 的相对含量。

Clark 氧电极就是一种氧分压传感器，可用于测量空气样本或血液样本中的氧气分压。它是一种安培传感器，使用时需要外加极化偏压电源。如图 9.19 所示，其测量原理基于极谱分析法，也就是在富含电子的情况下，氧分子 O_2 与水分子 H_2O 之间会产生化学反应，生成氢氧根离子 OH^-。这种电化学反应称为氧化还原反应，反应时需要外加 0.6V 左右的恒压极化电源，反应过程中会产生一个小电流。

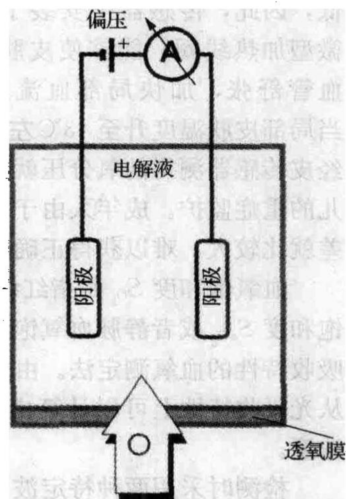
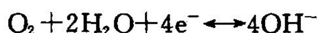


图 9.19　Clark 氧分压传感器的极谱分析原理

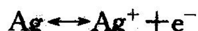
氧化还原反应中，氧气在铂或金等贵金属制成的

阴极电极（即连接电压源负极的电极）表面被不断还原，其化学反应式为

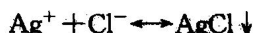


$$O_2+2H_2O+4e^- \longleftrightarrow 4OH^-$$

该还原反应中，每个氧分子获得 4 个电子，并与 2 个水分子起作用，生成 4 个氢氧根离子 OH^-。OH^- 离子流向 Ag/AgCl 参考电极，也就是与电压源正极相连的阳极，在那里发生以下两步氧化反应：



$$Ag \longleftrightarrow Ag^+ + e^-$$



$$Ag^+ + Cl^- \longleftrightarrow AgCl\downarrow$$

该氧化反应过程中，首先电极上的银被氧化成为银离子，电子则释放到阳极上；然后，银离子又立刻与氯离子结合，形成氯化银沉积在阳极表面。氧化还原反应在外电路的阴极与阳极之间产生的电流大小直接与阴极上不断还原的氧分子数目成正比。Clark 氧电极中，极谱池的电极浸在氯化钾电解液中，并用透氧聚四氟乙烯塑料薄膜或者聚丙烯薄膜将小池封住，使气体分子可以缓慢扩散到电极上。这样，只要测得阴极与阳极之间的电流，就可以计算出溶液中溶解的氧气量。

略微改变一下 Clark 极谱氧分压传感器的结构，就可以用它测量经皮氧分压。图 9.20 显示了这种经皮氧分压传感器的截面图，其实它是一个用双面胶贴在皮肤表面的标准 Clark 氧电极。与测量血样中的氧分压一样，它测量从皮肤下扩散出来的血氧分压。不过，皮肤的氧扩散率一般非常低，因此，传感器中安装了一个微型加热线圈，能够使皮肤毛细血管舒张，加快局部血流速度。当局部皮肤温度升至 43℃左右时，经皮传感器测得的氧分压就接近皮下动脉血的氧分压了。这种传感器广泛用于新生儿的重症监护。成年人由于皮肤较厚，皮肤气体扩散性质的个体差异很大，检测误差就比较大，难以获得正确的测量结果。

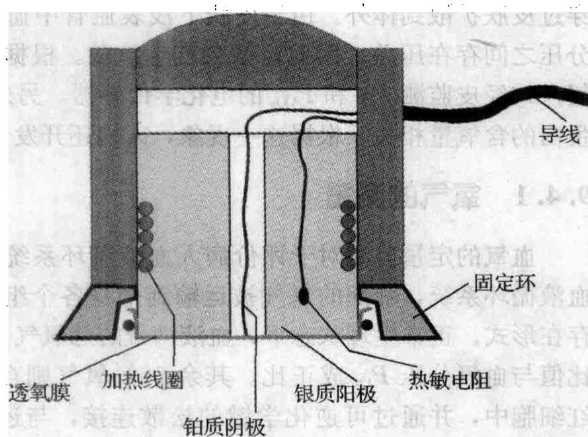


图 9.20 经皮氧分压传感器

血氧饱和度 S_{O_2} 是指红细胞血红蛋白的相对含氧量，离体或者在体检测动脉血氧饱和度 S_{aO_2} 或者静脉血氧饱和度 S_{vO_2} 的方法有很多种。这里将介绍的是基于血液光吸收特性的血氧测定法。由于无氧血液呈现蓝色，而含氧血液则呈现亮红色，因此，从光吸收特性上可以计算出血液中还原血红蛋白 Hb 和氧合血红蛋白 HbO_2 的相对浓度。

检测时采用两种特定波长的光：一是波长为 λ_1 的红光，如 660nm，此波长下两种血红蛋白 Hb 与 HbO_2 之间的吸光率差别很大；二是波长为 λ_2 的近红外光，可以选择 805nm 左右的波长，此波长下 Hb 与 HbO_2 的吸光率几乎相等。也可以选择 940～

960nm 的波长，此范围内 Hb 的吸光率略小于 HbO_2。图 9.21 显示了可见光和近红外区域血液的吸收光谱。

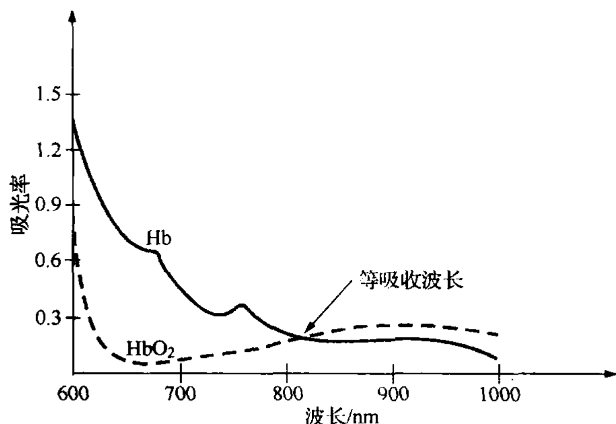


图 9.21　两种血红蛋白 Hb 和 HbO_2 的光吸收特性

这种检测方法的依据就是如下比尔-朗伯定律（Beer-Lambert's law）：

$$P_t = P_0 \times 10^{-abc} \tag{9.17}$$

式中　P_0——入射光功率；

P_t——透射光功率；

a——样品的消光系数，即摩尔吸收系数，是与波长相关的常数；

b——样品的光程长度；

c——样品浓度。

假设：（1）$\lambda_1=660$nm，$\lambda_2=805$nm（即等吸收波长）；（2）血样经过溶血处理，也就是红细胞已破碎，血红蛋白被释放并与血浆均匀混合，其中含有还原血红蛋白 Hb 和氧合血红蛋白 HbO_2；（3）两种蛋白的总吸收系数符合加法原理。那么，可以得到如下计算血氧饱和度的简单数学公式：

$$S_{O_2} = A - B \times \left[\frac{OD(\lambda_1)}{OD(\lambda_2)}\right] \tag{9.18}$$

式中　A 和 B——两个系数，分别是 Hb 和 HbO_2 各自的光吸收系数的函数；

OD——光密度，即吸光率，定义为 $\lg(1/T)$，其中 T 是样品的透光率，等于 P_t/P_0；

S_{O_2}——血氧饱和度，定义为 $C_{Hb}/(C_{Hb}+C_{HbO_2})$，其中 C 表示浓度。

血液 S_{O_2} 的检测有离体和在体两种方法。离体检测用台式光电血氧计，需要从外周动脉采集血样，然后将血样加入比色皿（即杯壁平行的玻璃样品杯），先溶血，再依次用两种波长的单色光进行测量。单色光是用窄带滤光片从强白光光源滤光获得。在体 S_{O_2} 检测使用非侵入式光电脉搏血氧计。例如，测量动脉血氧饱和度 S_{aO_2} 的脉搏血氧计由两个廉价的小发光二极管（LED）和一个高灵敏度的硅光电管构成。其中，

一个 LED 发射 660nm 左右的红光 R，另一个 LED 则发射 940～960nm 的红外光 IR。如图 9.22 所示，这几个元件一般安装在可重复使用的弹簧夹里，或者安装在一次性护套内。血氧计的电子电路会产生数字切换信号，控制两个 LED 交替点亮和熄灭，同时记录每个 LED 发光时的光电管输出信号。使用时，通常将血氧计的传感器夹在指尖或耳垂上，使被测人体组织处于光源与光电管之间。

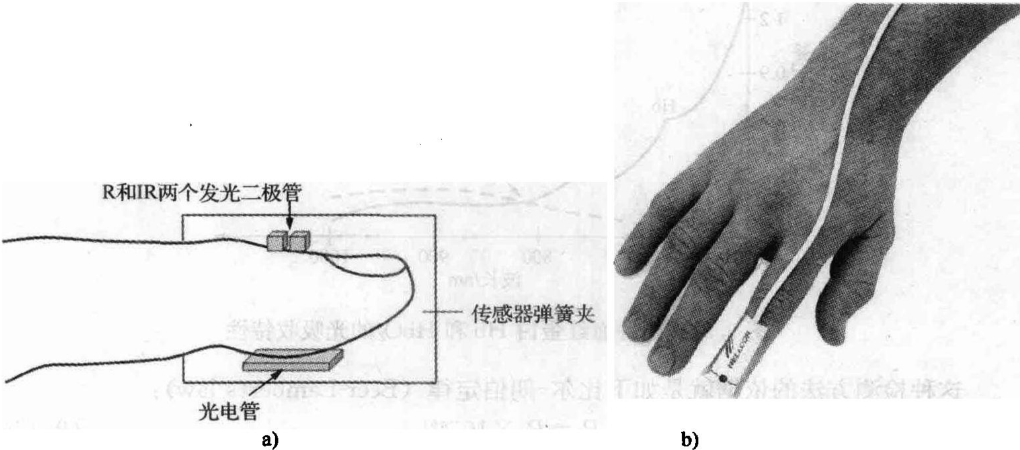


图 9.22 透射型光电式脉搏血氧计的指尖探头（图 a）和一次性指尖传感器（图 b）（由 Nellcor Puritan Bennett 公司（Pleasanton，CA）提供）

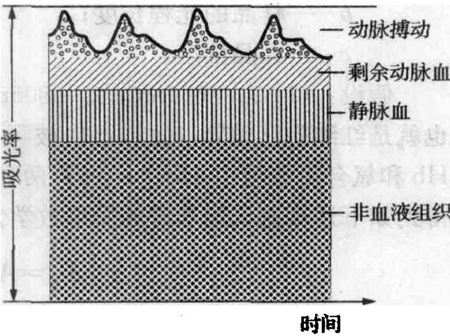


图 9.23 动脉搏动作用下外周组织血管网的光吸收系数随时间的变化曲线

如图 9.23 所示，脉搏血氧计检测到的是光信号的变化曲线，该信号反映了心脏周期性收缩引起的动脉血容量的变化。与信号幅值相关的参量有：每次心搏周期心脏射入外周血管网的血容量、血液的光吸收系数、皮肤和皮下组织的结构和颜色以及测量采用的照射光波长。通过分析光电管测得的光信号曲线中红光和红外光的幅值，可以求得 S_{aO_2}。电子电路可以将红光和红外光两种信号中所包含的脉动量（即交流分量 AC）和直流分量 DC 分开。其中，直流分量主要反映非血液组织、心脏舒张期剩余动脉血、静脉血以及皮肤色素的光吸收量，而交流分量则被认为仅仅与动脉血的搏动变化有关。血氧计用一种归一化算法，将每个波长下测得的 AC 信号除以 DC 信号，然后，再计算红光与红外光的比率 R。该 R 值与动脉血的颜色密切相关，也就是与 S_{aO_2} 相关，而与心脏搏动时射入人体组织的动脉血流量、皮肤色素和皮肤厚度等因素基本无关。因此，这种血氧计用于不同病人时无需校准。S_{aO_2} 与 R 之间对应的数学关系已由生产厂

家通过编程设置在仪器的程序中。

9.4.2　pH 电极

pH 值可以指示溶液的酸碱平衡状况。酸性溶液含有大量氢离子 H^+，而碱性溶液则含有大量氢氧根离子 OH^-。并且，在稀溶液中，这两种离子的浓度之积是常数，等于 1.0×10^{-14}。因此，两种离子中任意一个浓度都可以用来表示溶液的酸性或碱性。中性溶液的 pH 值等于 7.0。

血液 pH 值的检测是许多医学诊断的基础。人体正常血液的 pH 值受到严密的调控，一般保持在 7.40 左右，稍偏碱性。通过测量血液的 pH 值，可以判断肺部组织排除体内 CO_2 是否充分，还可以诊断肾脏调节酸碱平衡的功能是否正常。

pH 电极是一种能够产生微小电压差的传感器。如图 9.24 所示，它由独立的参比电极和测量电极构成，两个电极一般都用 Ag/AgCl 丝制作，并且浸在 KCl 电解液中，分别封装于各自的玻璃管内。参比电极是一个盐桥，其实就是装有电解液的一支玻璃管，其头部用允许通过各种离子的通透膜封住，这样，就可以将参比电极的电压保持在某个恒定值下，不会随被测溶液变化。而测量电极的头部则使用只能通透氢离子的玻璃膜封住。这两个电极还可以封装在同一个玻璃管护套内。

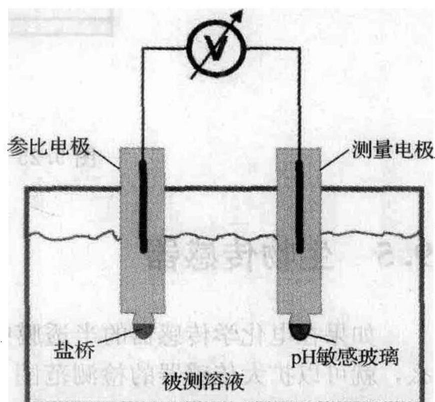


图 9.24　pH 电极的原理

在 KCl 溶液与被测溶液的接触界面上会产生一个正比于被测溶液氢离子浓度的电压差，25℃恒温下，此电压的计算公式是

$$V=-59\text{mV}\times\lg[H^+]+C \tag{9.19}$$

式中　C——常量。

由于 pH 的定义是

$$\text{pH}=-\lg[H^+] \tag{9.20}$$

因此，测量电极上的电压 V 与被测溶液的 pH 值成正比，即

$$V=59\text{mV}\times\text{pH}+C \tag{9.21}$$

常量 C 的值一般要用几种已知 pH 值的缓冲液进行标定，也就是将电极放入缓冲液，调节仪器进行补偿并校准。

9.4.3　CO_2 电极

图 9.25 所示是测量血液等液体中 CO_2 分压的电极，其原理与 pH 电极相同。因为 CO_2 溶于水后会形成弱电解质碳酸 H_2CO_3，使水中含有游离的氢离子和碳酸氢根离子，其化学反应式为



$$CO_2+H_2O\leftrightarrow H_2CO_3\leftrightarrow H^++HCO_3^-$$

从而使溶液的 pH 值发生变化，在 pH 测量电极与 Ag/AgCl 参比电极之间产生电压差，该电压与 CO_2 分压的负对数值成正比。

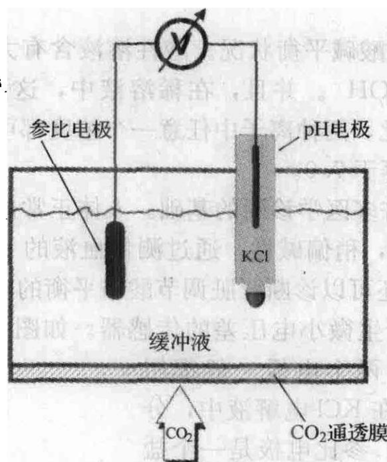


图 9.25 CO_2 分压检测电极的原理

9.5 生物传感器

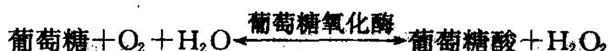
如果在电化学传感器的半透膜中加入反应物或者催化剂等生物特异性物质，那么，就可以扩大传感器的检测范围。如今，人们已经利用各种不同的酶和微生物作为主要传感元件，成功开发了多种传感器，不过，目前主要还是用于实验室研究。虽然这些传感器已成功地应用于离体检测，并且展现出独特的医学和工业领域的应用前景，但是，在技术上还需进一步改进，以提高其稳定性和可靠性，才能满足常规临床检验的要求。下面介绍几种主要生物传感器的设计原理。

9.5.1 酶传感器

酶是具有生物催化特性的一类蛋白质，大约包括 2000 多种蛋白质，它们在加速细胞化学反应方面具有独特且强大的功能。通常，酶具有很强的特异性，只对特定的底物产生作用，即使这些底物与其他物质混合在一起，酶也可以将它们识别出来。但是，必须注意，可溶性酶对于温度和 pH 值都非常敏感，并且可以被许多化学物质抑制，使酶失去活性。在实际生物传感器产品中，酶通常被固定在淀粉凝胶、硅橡胶、聚丙烯酰胺凝胶等稳定的惰性基质中，这样可以确保酶的催化活性，且可以重复使用。

利用不同的酶可以做成各种不同的传感器。例如，葡萄糖传感器是一种常用的酶传感器，它用了葡萄糖氧化酶。葡萄糖对于人体新陈代谢非常重要，糖尿病患者由于胰腺分泌的胰岛素不足，无法控制血液中的血糖含量，这时，就需要监测血糖

浓度，并通过注射胰岛素药物，将血糖浓度控制在正常范围内。目前所用的葡萄糖传感器采用的是葡萄糖氧化酶等固定化酶，通过电化学方法检测葡萄糖氧化过程中生成的葡萄糖酸、过氧化氢（H_2O_2）或者耗氧量，就可以计算出葡萄糖含量，其化学反应式为



$$葡萄糖 + O_2 + H_2O \xrightleftharpoons{葡萄糖氧化酶} 葡萄糖酸 + H_2O_2$$

如图 9.26 所示，葡萄糖传感器的结构与氧分压传感器很相似。葡萄糖和氧气穿过外层透氧薄膜进入传感器，并与葡萄糖氧化酶产生反应；反应剩余的氧气再穿过另一层透氧薄膜，到达内部的氧电极，从而测得氧浓度。

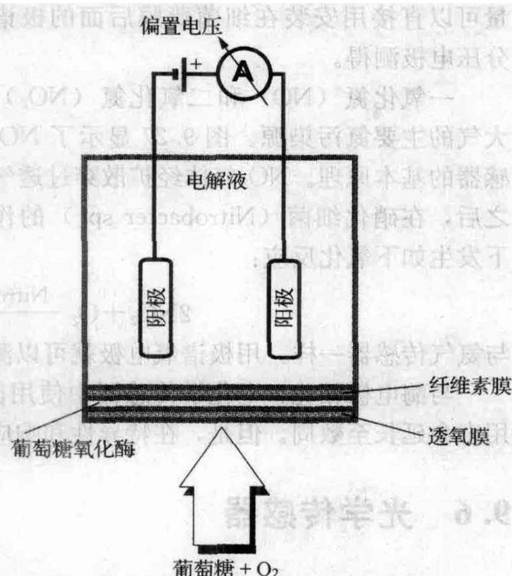


图 9.26　葡萄糖传感器的基本原理

生物酶传感器通常包含一个电化学气体敏感传感器，也就是离子选择电极，电极上附有一层固定化酶薄膜，作为生物催化剂。样品溶液中的被测物质经过扩散进入催化层，发生酶促反应，反应过程中产生或者消耗的电活性产物就由离子选择电极检测。如果反应中要消耗 O_2 等气体，或者会产生 CO_2、NH_3 等气体，那么，电极与酶之间还有一层薄膜隔开。显然，随着反应的进行，底物浓度会不断减小，但通常这种消耗可以忽略不计。只有当被测样品的量很少或者包含酶的膜面积很大时，才需考虑底物浓度的损耗。因此，通常可以认为这种电化学检测方法不会损坏被测样品，样品可以重复使用。检测时一般需要保持 pH 值和温度恒定不变，样品可以是搅拌均匀的溶液，也可以是处于流动状态的液体。

9.5.2　微生物传感器

微生物传感器产品已有许多种，主要用于环境工程、农业、食品业和医药等领域的在线生化反应控制。其检测基础是微生物同化有机化合物的反应，以及反应后的新陈代谢变化，也就是微生物分泌的 H_2、O_2、NH_3 等具有特定电化学活性的代谢产物的变化。

微生物传感器中含有固定化微生物，作为特异性识别元件；还有电化学传感器件或者光敏器件，用于将生化信号转化为电信号。传感器的工作过程包括如下 5 个步骤：将底物运送到传感器表面；底物经扩散穿过薄膜，到达固定化微生物；在微生物部位发生反应；反应产物穿过薄膜到达检测器探头；由探头测出产物含量。

例如，氨气（NH_3）和二氧化氮（NO_2）传感器就属于微生物传感器，两者都利用硝化细菌作为生物敏感元件。氨气传感器用亚硝化单胞细菌（Nitrosomonas sp.）等硝化细菌制作，硝化细菌能够将氨气作为能源，产生如下氨的氧化反应：



$$NH_3 + 1.5O_2 \xrightarrow{\text{Nitrosomonas sp.}} NO_2 + H_2O + H^+$$

该氧化反应的速度很快，固定化细菌的耗氧量可以直接用安装在细菌薄膜后面的极谱氧分压电极测得。

一氧化氮（NO）和二氧化氮（NO_2）是大气的主要氮污染源。图 9.27 显示了 NO_2 传感器的基本原理。NO_2 气体经扩散穿过透气膜之后，在硝化细菌（Nitrobacter sp.）的作用下发生如下氧化反应：

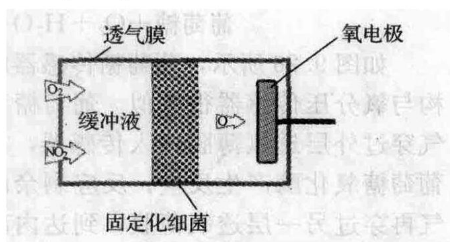
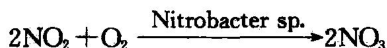


图 9.27　NO_2 微生物传感器的基本原理



$$2NO_2 + O_2 \xrightarrow{\text{Nitrobacter sp.}} 2NO_3$$

与氨气传感器一样，用极谱氧电极就可以测得反应膜周围的耗氧量。

与酶电极相比，电化学传感器中使用微生物细胞的主要优点是可以将电极的使用寿命延长至数周。但是，在特异性和响应时间上，微生物传感器就不如酶电极。

9.6　光学传感器

光学传感器对于开发高灵敏度、高选择性生化检测方法具有重要的作用。其基础是物质光特性的变化，包括吸光率、反射率、折射率、散射率、荧光特性、偏振特性等本征变化。

9.6.1　光纤

光纤可以将光从一个地方传导到另一个地方。如图 9.28 所示，光纤一般是透明玻璃或塑料制成的双层同心结构，内层称为纤芯，外层称为包层。

光纤纤芯和包层的折射率不一样。折射率等于光在真空中的传播速度与光在某种特定介质中的传播速度之比。例如，空气的折射率为 1.0，而水的折射率为 1.33。假设纤芯材料的折射率为 n_1，包层的折射率为 n_2，且 $n_1 > n_2$，那么，根据 Snell 定律有

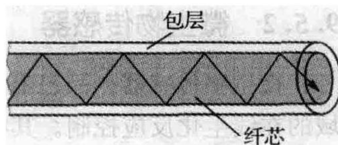


图 9.28　光纤的结构

$$n_1 \times \sin\phi_1 = n_2 \times \sin\phi_2 \qquad (9.22)$$

式中　ϕ_1——光的入射角（见图 9.29）；

　　　ϕ_2——光的折射角。

由此可见，当光线从折射率较小的介质射向折射率较高的介质时，其传播方向会向两种介质交界面的垂直方向偏转。如图 9.29 所示，在光纤的输入端，当相对于

芯轴的入射角 ϕ_1 较小时，光线会进入纤芯，并在第一次遇到纤芯与包层交界面时产生全反射；但是，当入射角 ϕ_1 较大时，如果此 ϕ_1 角超过了光线在纤芯与包层交界面上能够产生全反射所需的临界角，那么，光线就会逸漏到包层中去⊖。

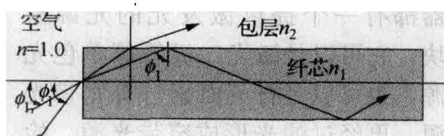


图 9.29 光纤中的入射光和反射光
注：实线表示从纤芯逸漏到包层的光线，虚线表示在纤芯内全反射的光线。

在纤芯与包层界面上，设 $\sin\phi_2=1.0$，可求得此处的全反射临界角 ϕ_{cr} 为

$$\sin\phi_{cr}=\frac{n_2}{n_1} \tag{9.23}$$

任何进入光纤的光线，如果在纤芯与包层界面上的入射角大于 ϕ_{cr}，都会被外层包层全反射回纤芯；反之，如果入射角小于 ϕ_{cr}，则光线会穿过包层逸漏出去，不能在纤芯中传播。

例 9.11 假设一束光线从折射率 $n_1=1.47$ 的玻璃层射入折射率 $n_2=1.44$ 的玻璃层，请应用 Snell 定律计算这两层玻璃交界面上的全反射临界角。

解：

$$\phi_{cr}=\arcsin\times\left(\frac{n_2}{n_1}\right)=\arcsin\ (0.9796)$$

$$\phi_{cr}=78.4°$$

因此，光线从第一层玻璃射向第二层玻璃时，如果其入射角大于 78.4°，就会被全反射回第一层玻璃。 ■

其实，即使光线在光纤中全反射传播，其传播也并不是仅仅局限于纤芯之内，它会穿过纤芯与包层的界面，进入光密度较小的包层内大约一个波长数量级的极小深度。这种效应会激发出一种与入射角和入射光波长相关的电磁场，被称为渐逝波(evanescent wave)。渐逝波起始于纤芯与包层的交界面，并延伸到包层内，其强度变化遵循比尔-朗伯定律，也就是随着深入包层的距离按指数衰减。因此，光线在光纤中传播的过程并不是严格在分界面上全部反射，而是穿透到包层内一定深度后逐渐全部反射。

9.6.2 光纤传感器的基本检测原理

检测生物传感器界面上光特性变化的光学方法主要有两种：一种是在光纤传感器中起重要作用的渐逝波光谱法，另一种则是表面等离子体共振法。

光纤可以用于开发各种各样生物医学传感器。这类传感器具有体积小、灵活、不受电气干扰等优点，并且能够对其周围微环境的变化产生即时响应。

早在 20 世纪末就已经成功开发用于血气监测的光纤传感器产品。虽然光纤传感器的种类很多，但它们都具有如图 9.30 所示的几个主要组成部分。首先，光纤传感

⊖ 光纤输入端光线的入射角与光线在纤芯一包层界面上的入射角成反比。——译者注

器都有一个提供激发光的光源模块，它可以是激光二极管等单色光源，也可以是石英卤素灯等宽带光源，再经过滤光形成窄带光源。检测时，通常要用到两个波长的单色光，其中一个波长的光对被测样品的浓度变化较敏感，而另一个波长的光则几乎不受被测样品浓度变化的影响。这样，用第二个波长的测量数据作为参照，可以补偿光源和检测器的不稳定因素引起的干扰。光源模块输出的光束需要经过一组透镜和光学接口之后，才能进入光纤，再由光纤传送给被测样品。

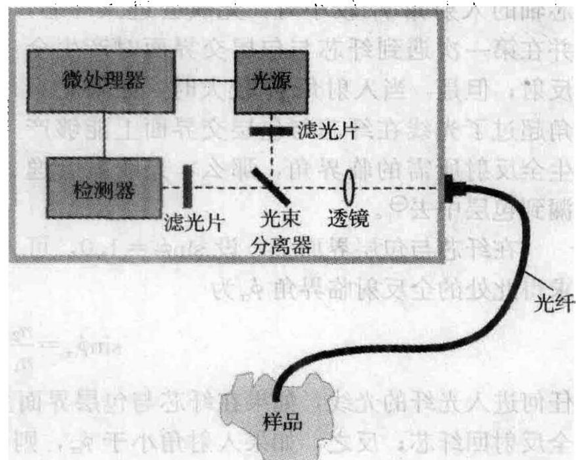


图 9.30 光纤传感器的基本组成结构

光纤传感器的检测部分可以根据多种不同的原理来设计。例如，在荧光式传感器中，由光纤传送的入射光激发被测溶液发射荧光，此荧光的强度与被测溶液的浓度之间存在一定的函数关系，当光纤将被测物发射的荧光传回检测器时，用光电探测器就可以检测出荧光的强度。另一类光纤传感器则是检测被测物的光吸收特性。这种吸收型传感器通过装在光纤头部的反射镜，或者由被测物本身的散射效应，将光返回到原光纤中传回。还有传感器利用渐逝波与其穿透深度范围内所含被测物分子之间的相互作用，这种作用会减小纤芯的光反射率，并且，反射率的减小与渐逝波穿透层内被测分子的浓度相关。由于渐逝波穿透层的深度很小，而且渐逝波的强度随距离按指数迅速衰减，因此，吸收渐逝波的被测物必须紧贴纤芯表面，光纤的包层要去掉。这种原理已经用于测定纤芯表面附着的受体与纤芯周围溶液中配体之间的化学反应。

渐逝波光纤检测法的关键是纤芯表面附着的某种生物介质层，其中的受体必须能够保持其原有的结构和活性，并且要有足够的结合位点能与被测样品中的配体相结合。如果被测物的光吸收性很弱，那么，就需要结合多重内反射，也就是沿着除去包层的光纤，光线要发生多次内反射，这样，就可以提高检测的灵敏度。另外，除了利用物质的光吸收特性以外，还可以利用荧光特性，也就是检测在光照作用下会发射荧光的物质。荧光物质吸收激发光以后，会发射出波长比激发光长的可测荧光，这样也能提高检测的灵敏度。

9.6.3 使用指示剂的光纤传感器

通过标准光谱法就能够直接检测到本征光吸收和荧光发射的生化物质很有限，只有少数几种，因此，人们开发了利用指示剂作为中介物的光纤传感器。这种传感器的光纤表面或者头部附近固定有特定的试剂，当光源发出的光束传播到光纤末端

时，就会与这些特定的试剂（即生物识别元素）发生作用。试剂可以是指示剂，离子键合物（即离子载体）以及各种具有选择性的聚合物。入射光与被测样品作用之后，返回的光通过原光纤（单光纤结构）或者通过另一根独立的光纤（双光纤结构）回到探测器，再由探测器检测出光的衰减程度，并由此计算被测样品的浓度。

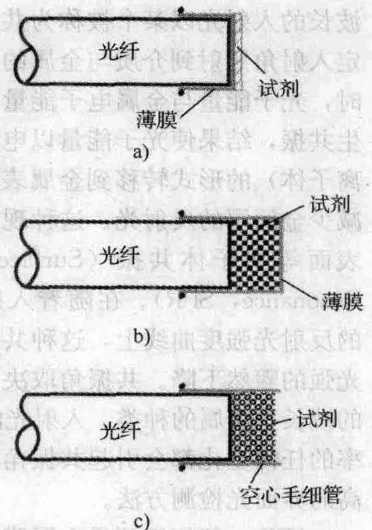


图 9.31　几种使用指示剂的光纤传感器的结构

图 9.31 显示了几种指示剂中介型传感器的典型结构。传感器的检测部分就是光纤头部的一层化学试剂，它与被测溶液之间用一层选择性通透膜隔开。化学试剂能够将入射光转换成回传的光信号，此回传光的强度与被测物浓度成正比。传感器的稳定性与所用试剂的光敏特性的稳定性相关，也与探头上试剂封装的可靠性相关，试剂不能被浸出。如图 9.31a 所示的结构中，试剂被直接固定在光纤头部的薄膜上。如果试剂是一种粉末，那么，可以用特殊的通透膜（见图 9.31b）或者空心毛细管（见图 9.31c）将其封装在光纤头部。

9.6.4　免疫传感器

免疫传感器设计的基础是被测样品与高特异性试剂之间存在配位键反应。这种传感器的关键是生物识别元件，一般是抗体或者抗体片断。抗体与抗原相结合的免疫识别机制使得免疫传感器具有极好的选择性，能够从含有其他物质的复杂混合物中检测出特异性物质，并且灵敏度极高，能够从样品中检测出微摩尔甚至皮摩尔级浓度的重要化合物。

渐逝波免疫传感器可以用于免疫诊断，检测抗体与抗原的结合状况。图 9.32 所示是这种传感器的原理图。被测样品通常被注入一个流动小室，小室内装有光纤探头，抗体就固定在除去包层的光纤头部表面，这些抗体能够捕获样品溶液中的抗原。当抗体与抗原反应结束之后，小室内的样品溶液被换成带标记的抗体。这些带标记的抗体与已经被捕获在光纤头部的抗原结合之后，就会被激发出荧光。传感器检测此荧光信号，并由此计算出样品溶液的抗原浓度。

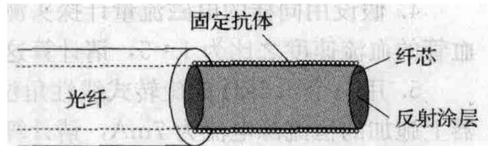


图 9.32　光纤免疫传感器的原理

9.6.5　表面等离子体共振传感器

如果将一束单色偏振光（如激光）照射到表面涂有金属导体（如银、金等）的透明介质上，那么，界面处就会产生一种电荷密度振荡。如图 9.33 所示，如果适当

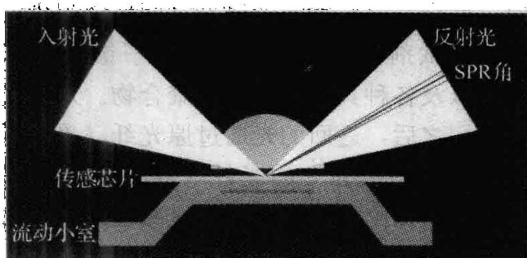


图 9.33 表面等离子体共振 SPR 的检测原理（由瑞典乌普萨拉市 Biacore AB 馈赠）

波长的入射光以某个被称为共振角的特定入射角照射到介质与金属的交界面上时，光子能量与金属电子能量之间会发生共振，结果使光子能量以电子包（即离子体）的形式转移到金属表面，同时减少金属层的反射光。这种现象被称为表面等离子体共振（Surface Plasmon Resonance，SPR）。在随着入射角变化的反射光强度曲线上，这种共振表现为光强的骤然下降。共振角取决于入射光的波长、金属的种类、入射光的偏振状态以及金属膜下面的介质的特性。介质折射率的任何变化都会引起共振角的变化，因此，根据这个原理可以设计一种灵敏度极高的界面光检测方法。

SPR 一般用于测量金属膜下介质折射率的微小变化。例如，在其他参数都保持不变的情况下，即使金属表面只粘附了一个抗体分子，其折射率的变化就足以引起可测量的共振角的明显变化。这种方法的优点是能够很好地检测界面上抗体与抗原的反应。

习　　题

1. 将两个相同的银电极插入电解质溶液中，请计算两电极之间的电压差。

2. 将镉电极和银电极插入电解质溶液中，假设溶液的等效电阻为 8kΩ，请计算流过两个电极的电流。

3. 如果将电感式位移传感器的线圈匝数增加 4 倍，那么线圈的电感将变为多少？

4. 假设用同样的电磁流量计探头测得两条血管的感应电压之比为 2∶3，且两条血管的血流速度之比为 1∶5，请计算这两条血管的截面积之比。

5. 用一个 4.5kΩ 的旋转式线性角位移传感器测量膝关节的角度变化，假设传感器上施加的恒流源电流为 7mA，请计算膝关节旋转 165°时传感器的输出电压。

6. 请写出式（9.11）的推导过程。

7. 某弹性电阻式传感器的初始电阻为 R_0，初始长度为 l_0，将其拉伸。假设拉伸过程中传感器的截面积产生变化，请推导传感器电阻变化量 ΔR 的数学计算公式，并用 l_0、长度变化量 Δl、体积 V 和电阻率 ρ 表示。

8. 差动式电容传感器的各块电容极板面积均为 4cm^2，假设传感器处于平衡位置时，两个电容的极板间距均为 2mm，请计算板间介质为空气时各电容的平衡值。

9. 请作出电容传感器的电容值（y 轴）与位移（x 轴）之间的关系曲线图。

10. 请计算小位移变化下电容传感器的灵敏度，即 $\Delta C/\Delta d$。

11. 将电容传感器安装在床垫中，用于监测婴儿的呼吸状况。假设呼吸过程中，

电容的电压变化率（即 dV/dt）为 $\pm 1V/s$，并且电压变化曲线可以用三角波来模拟。请作出该传感器的电流变化曲线图。

12. 请推导压电晶体等效电容中流过的电流与电压 V 和电容 C 之间的关系式。

13. 如图 9.34 所示，将两个相同的超声波传感器放在血管两侧，如果超声波从一个传感器传到另一个传感器所需的时间为 250ns，请计算血管的直径。

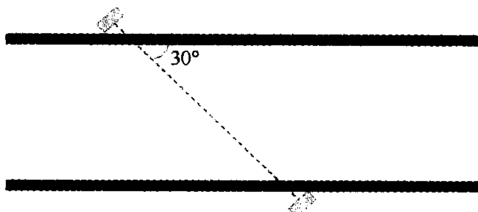


图 9.34　两个相同的超声波传感器放在血管两侧

14. 假设某热敏电阻在 12℃时的电阻值为 3.5kΩ，并且 $\beta=4600$，请计算其在 98℉时的电阻值。

15. 某个 $\beta=5500$ 的热敏电阻在 18℃时测得的电阻值为 250kΩ，请问当其电阻值增加 1 倍时的温度是多少？

16. 已知某热敏电阻在 21℃室温时的电阻值为 3.2kΩ，当室温上升 10%时，其电阻值为 1.85kΩ。请计算该热敏电阻的 β 值。

17. 请作出典型 Clark 极谱氧电极的电流（y 轴）与氧分压（x 轴）的关系图。

18. 请解释为什么脉搏血氧计的归一化比率 R 的值与心脏收缩时进入人体组织的血流量大小无关。

19. 请解释为什么脉搏血氧计的归一化比率 R 的值与皮肤色素无关。

20. 请说明电压传感器和电流传感器之间的区别。

21. 将某 pH 电极连接在高灵敏度的电压计上，当电极浸入 pH 值为 5.25 的缓冲液时，电压计的读数为 0.652V；再将电极移到另一种缓冲液时，电压计读数降低了 25%。请计算第二种缓冲液的 pH 值。

22. 请作出某溶液的光密度 OD（y 轴）与溶液浓度（x 轴）的关系图，并求其斜率。

23. 某样品溶液的浓度为 1.55×10^{-3}g/L，将其置于光程为 1cm 的比色皿中，其透光率为 44%。改变样品溶液的浓度后，其透光率上升到 57%，求此时溶液的浓度。

24. 如果有一束光线从空气入射到水面，入射光与界面法线成 62°角，请计算其在水中的折射角。

25. 请说明为什么光纤传感器一般都需要同时进行双波长测量。

26. 如图 9.35 所示，利用化学传感器检测的吸收光谱测量某种染料的 pH 值。假设染料在各种酸碱性条件下的光吸收系数与其 pH 值呈线性关系。请设计一种计算染料 pH 值的方法。

27. 请说明吸收光谱与荧光光谱这两种检测方法的区别。

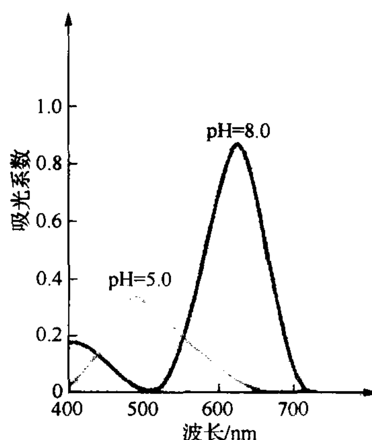


图 9.35 某种染料在酸性（pH=5.0）和碱性（pH=8.0）下的吸收光谱

推荐阅读资料

Allocca, J.A. and Stuart A. (1984). *Transducers: Theory and Applications*. Reston Publishing, Reston, VA.

Aston R. (1990). *Principles of Biomedical Instrumentation and Measurement*. MacMillan, New York.

Buerk, D. (1995). *Biosensors: Theory and Applications*. CRC, Boca Raton, FL.

Cobbold R.S.C. (1974). *Transducers for Biomedical Measurement: Principles and Applications*. Wiley, New York.

Cooper, J. and Cass A.E.G. (2004). *Biosensors*, Oxford University Press.

Cromwell, L., Weibell. F.J. and Pfeiffer, E.J. (1980). *Biomedical Instrumentation and Measurements*. Prentice Hall, Englewood Cliffs, NJ.

Eggins, B. (1997). *Biosensors: An Introduction*. Wiley, New York.

Eggins, B.R. (2002). *Chemical Sensors and Biosensors for Medical and Biological Applications*. John Wiley, New York.

Geddes L.A. and Baker L.E. (1989). *Principles of Applied Biomedical Instrumentation*, 3rd Ed. Wiley-Interscience, New York.

Hall, E.A.H. (1991). *Biosensors*. Prentice Hall, Englewood Cliffs.

Harsanyi, G. (2000). *Sensors in Biomedical Applications: Fundamental Technology and Applications*. CRC, Boca Raton, FL.

Neuman M.R. (1999). Biomedical Sensors. In *The Biomedical Engineering Handbook*, 2nd Ed. (J. D. Bronzino, Ed.). CRC/IEEE, Boca Raton, FL.

Togawa T., Tamura T. and Oberg P.A. (1997). *Biomedical Transducers and Instruments*. CRC, Boca Raton, FL.

Webster, J.G. (1988). *Encyclopedia of Medical Devices and Instrumentation. John* Wiley, New York.

Webster, J.G. (1998). *Medical Instrumentation: Application and Design*, 3rd Ed. John Wiley, New York.

Wise, D.L. (1991). *Bioinstrumentation and Biosensors*. Marcel Dekker, New York.

第10章　生物信号处理

Monty A. Escabí 博士[⊖]

本章目录

⊖ Susan Blanchard、Carol Lucas和Melanie T. Young也参与了本章的撰写工作。——原文注

本章主要内容：

- 各种生物信号的不同来源；
- 周期信号与暂态信号、确定性信号与随机信号之间的区别；
- 模/数转换；
- 采样定理；
- 傅里叶变换及其应用；
- Z变换；
- 线性系统的基本性质；
- 滤波和信号平均的基本原理；
- 模糊逻辑的基本原理和特点；
- 人工神经网络的基本原理。

10.1 绪论

生物的电、化学和力学活动往往会产生可测量和分析的信号，就是各种生物信号，如心脏搏动、肌肉收缩等活动在时间和空间上的记录信号。生物信号包含了非常有用的信息，可以用于了解特定生物现象或生物系统的生理机制，也可以用于医学诊断。

生物信号的采集方法有很多种。例如，医生用简单的听诊器就可以听到病人的心音。如果借助先进的高新技术生物医学仪器，那么就可以获得更多生物信号。采集到数据之后，为了提取其中的有用信息，要对生物信号进行分析。最基本的信号分析方法有放大、滤波、数字化和存储等，它们通常可以用简单的电子电路或者计算机来实现。除这些基本方法以外，常用的方法还有信号平均、小波分析和人工智能等高级数字信号处理方法，它们对于改善生物信号的质量具有重要的作用。

10.2 生物信号的生理基础

10.2.1 生物电信号

神经细胞和肌肉细胞会产生生物电信号，这些电信号是细胞内和细胞之间电化学变化引起的（详见第 11 章）。如果神经细胞或者肌肉细胞受到足够强的刺激，达到了所需的阈值，那么，细胞就会产生动作电位。动作电位由短促的细胞膜跨膜离子流形成，可以用胞内或胞外电极检测。神经细胞产生的动作电位可以通过轴突从一个细胞传导到其他细胞。大量细胞同时产生动作电位时会在细胞外形成一个电场，此电场可以通过生物组织传播出去。因此，这种胞外的电位变化可以在组织表面或者生物体表面用表面电极测量。例如，心电图（ECG）、胃电图（EGG）、脑电图（EEG）和肌电图（EMG）等都是这类生物电信号（见图 10.1）。

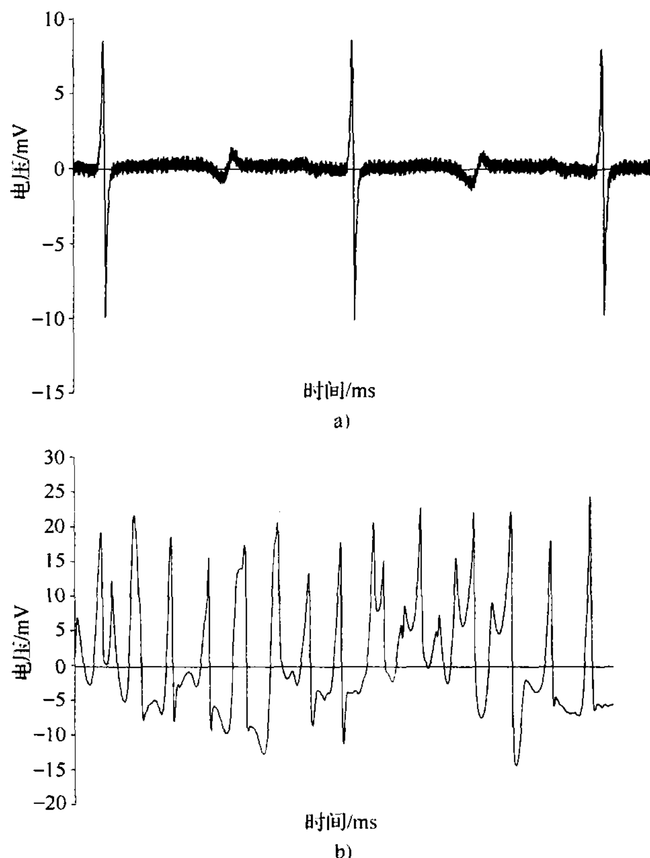


图 10.1 某头猪心脏表面的正常窦性心律心电图记录（图 a）和该猪心脏表面记录的心室颤动（Ventricular Fibrillation，VF）心电图（图 b）（采样频率为 1000 Hz）。

10.2.2 生物磁信号

心脏、大脑和肺等各种器官也会产生微弱的磁场，可以用磁场传感器测量。一般，磁场强度比相应的生物电信号要弱得多。生物磁学（biomagnetism）研究的就是与特定生理活动相关的磁信号，这种磁信号一般都与特定的组织或器官所产生的电场具有密切的联系。利用非常精密的磁场传感器，如超导量子干涉计（Superconducting Quantum Interference Device，SQUID）就可以直接监测大脑、外周神经、胃肠道和心脏等的磁活动，这些测量分别称为脑磁图（magnetoencephalography，MEG）、神经磁图（magnetoneurography，MNG），胃肠磁图（gastrointestinal tract，MGG）和心磁图（magnetocardiography，MCG）。

10.2.3 生物化学信号

生物化学信号反映了体内各种化学物质浓度变化的信息。例如，细胞内钙、钾等各种离子的浓度都可以检测和记录；呼吸系统或者血液中的氧分压（P_{O_2}）和二氧化碳分压（P_{CO_2}）等测量数据常用于检查血氧浓度是否正常，这些都是生物化学信号。生化信号有很多用途，比如通过测定血糖、乳酸和各种代谢物的水平，就可以了解各种生理系统的状况。

10.2.4 生物力学信号

生物系统运动所产生的位移、张力、压强和流量等力学参量也是生物信号，如血压就是血液对血管壁的作用力的测量。如图 10.2 所示，血压的变化可以用记录曲线表示。曲线的上升相表示心室收缩期，波峰表示血液从心脏射出，流向人体其他

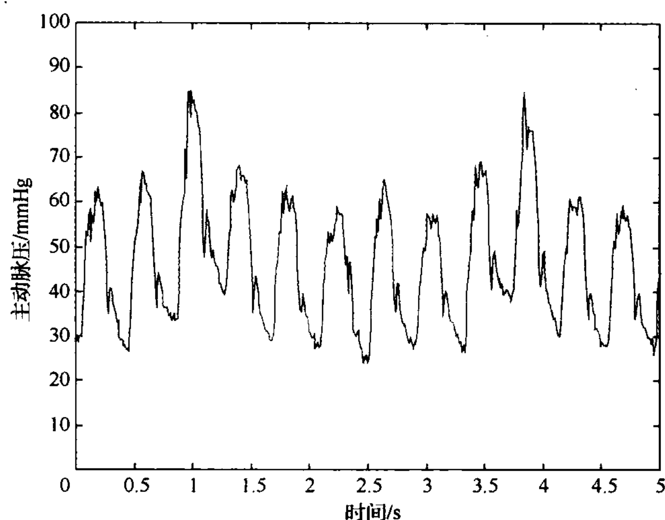


图 10.2 某 4 岁儿童的主动脉弓血压波形（采样频率为 200Hz）

部位，此时的血压称为收缩压，它是血压的最大值（详见第 3 章）；曲线的下降相则对应于心室舒张期，血压最小值称为舒张压。

10.2.5 生物声信号

生物声信号是涉及振动的一种特殊生物力学信号。很多生物现象都会产生声音。例如，血液流过心脏瓣膜时有一种独特的声音，通过测量这种心脏瓣膜所产生的声音信号，就可以判断心脏的功能是否正常。此外，呼吸系统、关节和肌肉也会产生声音信号，并且这些信号都可以经过生物组织传播出来，通常可以用放在皮肤表面的扩音器或者机械振动仪等声音传感器将它们检测出来。

10.2.6 生物光信号

生物光信号是由生物系统的光学特性或者光诱导特性产生的，也就是生物光信号可以是自发产生的，也可能是为了测量某种生物参量由外界光介质的引入而产生的。例如，通过测量羊水的荧光特性，可以判断胎儿的健康状况；用染料稀释法，也就是监测染料进入血液循环系统后的浓度，可以估计心输出量。另外，红光和红外光有各种各样的应用，例如，通过检测皮肤等特定组织对光的吸收量可以精确测量组织的血氧浓度。

例 10.1 当你快速跑过人行道时，你的小腿肌肉会产生哪些生物信号?

解：脚踩在路面上时，肌肉的运动和外力的作用会产生生物力学信号；神经对肌肉的刺激和肌细胞的收缩会产生生物电信号；肌肉组织的代谢过程则会产生生物化学信号。 ■

10.3 生物信号的特性

生物信号可以按照波形、统计特性和时域特性的不同进行分类。最常见的是分成连续信号（continuous signal）和离散信号（discretesignal）两大类。连续信号定义在连续的时间或空间上，用连续变量的函数来描述。符号 $x(t)$ 用于表示随连续时间 t 变化的连续信号 x。生物现象所产生的信号几乎都是连续信号，例如，图 10.1 所示的心电信号、图 10.2 所示的动脉压以及脑电信号等都是连续信号。

离散信号也是如今临床应用中常见的一类信号。它与连续信号不同，连续信号定义在连续的时间或空间上；而离散信号只定义在按一定规律分布的一部分时间或空间点上。因此，离散信号可以用数组，也就是数的序列来表示。符号 $x(n)$ 用于表示只在离散时间点 n 上存在的离散序列 x，其中 $n=0$，1，2，3，...，它总是一个整数，表示离散序列的第 n 个元素。虽然多数生物信号本身并不是离散信号，但是，如今数字技术迅速发展，离散信号具有非常重要的作用。高级医学仪器一般都将人体的连续信号转变成计算机可以分析和处理的离散数字序列。例如，从不同观察角度摄取病人的 X 射线图像（连续信号）之后，CT 扫描机的计算机断层成像技术就可

以从中提取数字化的采样图像数据，然后，对这些离散化的断层图像进行增强、重建等各种数字信号处理，形成完整的病人体内器官的三维计算机模型（详见第 15 章）。这种技术已经成为必不可少的临床诊断工具。

生物信号还可以分成确定性信号（deterministic signal）和随机信号两大类。确定性信号可以用数学函数来描述。周期信号和暂态信号就属于确定性信号。周期信号通常由不同的正弦波分量构成（见本章 10.5.1～10.5.3 节），可以表示为

$$x(t)=x(t+kT) \tag{10.1}$$

式中 $x(t)$——信号；

k——整数；

T——周期。

周期 T 表示周期信号上前后两个相邻重复波形在时间轴上的距离。周期信号其实是持续时间为 T 的基本波形的无限重复信号。暂态信号只在有限的一段时间内有不等于零的变化值，以后，随着时间的推移就逐渐衰减为某个常数值。图 10.3a 所示的正弦波就是周期信号的简单例子，它以 1s 为间隔不断重复出现。图 10.3b 所示是指数衰减信号与正弦信号的乘积，它是一个暂态信号，随着时间的推移，其信号幅值逐渐趋于零。

实际生物信号几乎总是含有不可预测的噪声或变化，因此不完全是确定性信号。例如，处于安静状态的正常心脏的 ECG 看上去几乎是周期性信号，但其实稍微含有一些不可预测的成分。ECG 的基本波形由 P 波、QRS 复合波和 T 波组成，并不断重复出现（见图 3.22）。但是，每一次心跳周期中 P 波、QRS 复合波和 T 波都会有一些不规则的变化。QRS 复合波之间的时间间隔，也就是常说的 R－R 间期，也会随时间发生变化，从而产生心率变异（Heart Rate Variability，HRV）。HRV 是一个诊断指标，可以用于心脏病发作之后的预后诊断，HRV 小的病人的预后往往比 HRV 大的病人差。

随机信号的参数中包含了不确定因素。由于这种不确定性，随机信号不能用数学函数精确表述，而要用统计学方法分析，也就是用信号的概率分布或者均值和标准差等统计量来描述。例如，EMG 就是一种随机信号，它是骨骼肌电活动的记录信号，可以用于神经肌肉疾病的诊断。稳态随机信号（stationary random signal）的统计特性和频谱特性不随时间变化；相反，非稳态随机信号的统计特性和频谱则随时间变化。在很多情况下，识别随机信号中的稳态部分对于信号处理、模式识别和临床诊断具有重要意义。

例 10.2 心室颤动（VF）是一种心率失常，它表现为 QRS 复合波和 T 波的无规律出现，也就是无规律的心肌收缩（见图 10.1b）。VF 常会导致猝死，在美国 VF 是死亡率最高的疾病之一。请问 VF 发生期间，ECG 记录信号最可能呈现出哪种类型的生物信号？

解： 心室颤动时 ECG 记录信号会是随机的连续生物电信号。 ■

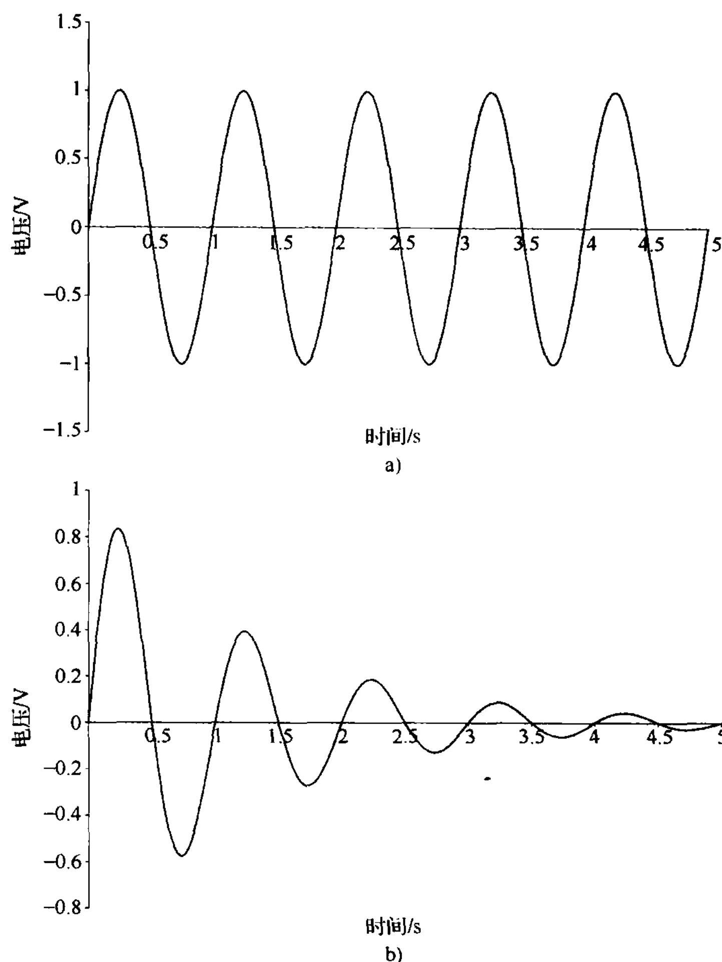


图 10.3　周期为 1s 的正弦波信号 $x(t)=\sin(\omega t)$（图 a）
和暂态信号 $y(t)=e^{-0.75t}\sin(\omega t)$（其中包含的正弦波周期也是 1s）（图 b）

10.4　信号采集

10.4.1　生物信号数据采集概述

生物信号通常很微弱，且一般都含有噪声信号。噪声会干扰测量信号中的有用信息，它可以是外来的，来自体外噪声源，如传感器的热噪声、数据采集系统电子装置中由供电系统引起的 60Hz 噪声等。噪声也可以是生物体内部产生的，它们来自测量部位附近的组织和器官。例如，心脏的 ECG 测量会受到周围肌肉组织电活动的影响。

为了能够从记录信号中提取对于了解特定生物系统或生物现象至关重要的信息，通常必须使用先进的数据采集技术和仪器设备，用高精度低噪声的仪器尽可能减小噪声的干扰。图 10.4 所示是生物仪器系统的基本结构示意图。

数据采集的整个过程中关键是要真实地保留感兴趣的原始生物信号的信息和结构，因为这些信号将用于辅助疾病的诊断，应尽可能避免失真。信号的放大、模拟滤波以及 A/D 转换等处理过程都不应该产生容易误解的失真或者难以理解的失真。测量信号的失真会导致误诊。

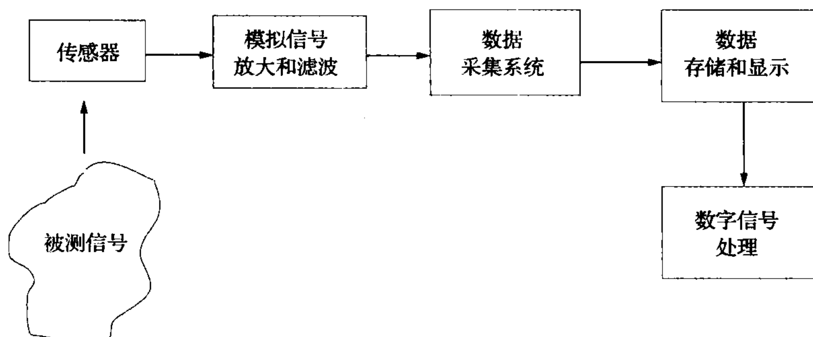


图 10.4 传感器将被测信号转换成可以测量的模拟电信号，数据采集系统再将模拟信号转化为可以储存的数字信号，数字信号处理技术则用于去除记录信号中的噪声，并提取原始信号中能够反映特定生理含义的有用信息

10.4.2 传感器、放大器和模拟滤波器

无论是细胞内部的生物信号还是皮肤表面的生物信号，都要用传感器检测（详见第 9 章）。传感器可以将物理量转换为电量输出，它是生物系统与电信号记录仪器之间的接口。生物信号的类型决定了所要使用的传感器的类型。例如，ECG 用粘贴在体表的电极测量，这种电极的银/氯化银（Ag/AgCl）界面可以检测离子的活动；而动脉压则用压力传感器测量。传感器检测的关键是传感器本身不能反过来影响被测生物信号的特性。

传感器测得的生物信号通常要经过放大和滤波。实际使用的放大器都是电子电路，其主要功能是增加生物信号的幅值。例如，生物电信号通常很微弱，有时需要放大上千倍，然后再用模拟滤波器去除噪声或者补偿传感器引起的失真。另外，为了满足数据采集系统的硬件要求，也必须进行生物信号的放大和滤波，因为只有将连续信号限制在一定的频率范围之内，才可以用模/数转换器进行数字化采样并将数据存储到计算机中。

10.4.3 模/数转换

模/数（Analog—to—Digital，A/D）转换用于将连续的模拟信号转换为数字序列。A/D 转换器其实是一种计算机控制的电压表，它测量模拟输入信号并给出数字

形式的输出信号。图 10.5a 所示是一个模拟信号，图 10.5b 则是其相应的数字序列。由传感器检测并经过放大和滤波的模拟信号是一种连续波形，A/D 转换器将这种连续的模拟信号转换成为离散的数字信号。离散信号由数字序列组成，很容易用数字计算机存储和处理。A/D 转换特别重要，因为随着计算机技术的发展，生物信号的存储和分析越来越依赖于计算机。

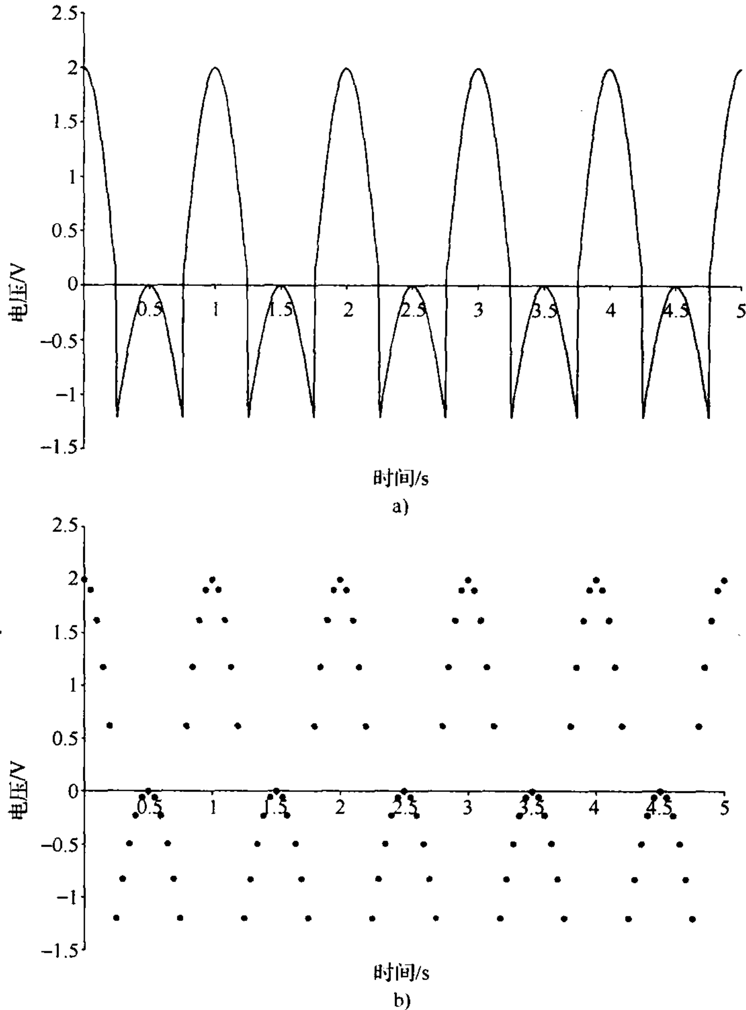


图 10.5 连续的模拟形式周期信号（图 a）和该模拟信号的数字形式（图 b）

经过 A/D 转换之后，原始模拟信号就可以用一系列数字信号来表示。但是，数字化信号并不能完全重现原始信号，离散数字信号只是原始模拟信号在固定时间间隔上的幅值的采样，是原始信号的数字化近似值。

A/D 转换包括采样和量化两个主要过程。采样是第一步，它把连续信号转换成

时间上离散的序列。假设 $x(t)$ 为模拟信号，采样就是每隔 T 秒时间记录一个 $x(t)$ 幅值。该幅值用 $x(kT)$ 表示，其中 $k=0, 1, 2, 3, \cdots$，为整数，表示采样数据序列中数据的位置；T 表示采样间隔，也就是两个相邻采样点之间的时间。实际应用中，数字信号处理一般只用有限长数据序列，因此，数据点的范围为 $k=0, 1, \cdots, N-1$，其中 N 是离散采样点的总数。采样频率 f_s 等于采样周期的倒数，即 $1/T$，其单位是赫兹 Hz，即 s^{-1}。

以下是几个特别重要的离散数字序列：

1）单位冲激脉冲序列： $\delta(k)=\begin{cases}1 & k=0\\ 0 & k\neq 0\end{cases}$

2）单位阶跃序列： $\mathrm{u}(k)=\begin{cases}1 & k\geqslant 0\\ 0 & k<0\end{cases}$

3）指数序列： $a^k u(k)=\begin{cases}a^k & k\geqslant 0\\ 0 & k<0\end{cases}$

要使离散的采样结果能够精确逼近连续信号，必须选择合适的采样频率。如果采样频率太低，数字信号就会产生失真。奈奎斯特定理（Nyquist's theorem）规定，如果要保留模拟信号的全部信息，那么，最小采样频率至少应该是原始信号所含最大频率的两倍。即 Nyquist 采样频率为

$$f_{\mathrm{nyquist}}=2f_{\max} \tag{10.2}$$

式中 $f_{\max}$——模拟信号中含有的最高频率成分。

因此，采样频率 f_s 必须大于等于 $2f_{\max}$，才能用数字序列完整表达模拟信号。实际使用时，采样频率通常为最高频率 $f_{\max}$ 的 5～10 倍。

A/D 转换的第二步是信号的量化（即数字化）。量化是将离散信号的模拟量幅值转化为计算机数字量。理论上，任意连续信号的幅值都要用无限长的数字才能表示。而计算机内存的容量有限，不可能存储任意数值。量化就是把信号幅值的取值限定在计算机可以处理的有限个数值上。

计算机里存储和处理的数字化数据通常都用二进制数字来表示，因此，采样过程中所产生的每个采样值都必须量化，也就是，离散的采样序列都要转化为二进制数字。A/D 转换器的分辨率决定了采样数据的位数。一般，A/D 转换器有 8 位、12 位或 16 位（即二进制位数）等分辨率。如果位数不够多，那么数字化逼近就可能出现明显的误差。采样所得二进制数值与原始数值之差称为量化误差。

A/D 转换器能够产生的二进制位数是其重要性能指标。N 位 A/D 转换器可以表示 2^N 个可能的幅值。二进制位数越多，A/D 转换器的分辨率就越高。16 位 A/D 转换器能够表示 65536 个不同的幅值，其分辨率高于只能表示 256 个幅值的 8 位 A/D 转换器。用模拟输入信号的电压范围除以 A/D 转换器可能取值的个数就可以计算出 A/D 转换器的分辨率。

例 10.3 假设输入信号的电压范围为 10V，请计算 8 位 A/D 转换器的分辨率。

解：
$$\frac{\text{输入电压范围}}{2^N}=\frac{10\text{V}}{256}=0.0391\text{V}=39.1\text{mV}$$
■

例 10.4 EEG 模拟信号的频率范围是 0.5～100Hz。要获得正确的数字信号，则其采样频率至少为多少？

解： 模拟信号的最高频率＝100Hz，

所以
$$f_{\text{nyquist}}=2f_{\max}=2\times100\text{Hz}=200\ \text{采样数/s}$$
■

那么，如果信号的采样频率不够高，会产生怎样的结果呢？采样定理的另一种表述方式是：设采样频率 $f_s=1/T$，那么，采样之后，信号中所有 $[f-kf_s]$ 频率上的值就都相同，其中 $-\infty\leqslant k\leqslant\infty$。由此不难看出上述问题的答案。

例 10.5 假设某个 360Hz 信号的采样频率为 200Hz，那么，采样得到的数字信号的频率会是多少？

解： 这里 $f_s=200\text{Hz}$，因此，采样之后，数值都相同的频率应该是 $[360-k200]=[\cdots360\ 160\ -40\ -240\cdots]$。在这组信号中只有 40Hz 的信号是 200Hz 采样频率可以正确表示的，因为此时采样频率大于信号频率的两倍。对于实数信号，－40Hz 和 40Hz 的信号看起来是一样的，即 $\cos(-\omega t)=\cos(\omega t)$，$\sin(-\omega t)=-\sin(\omega t)$。因此，本题中采样得到的信号看起来会是 40Hz 的信号。图 10.6 显示了这个采样过程。 ■

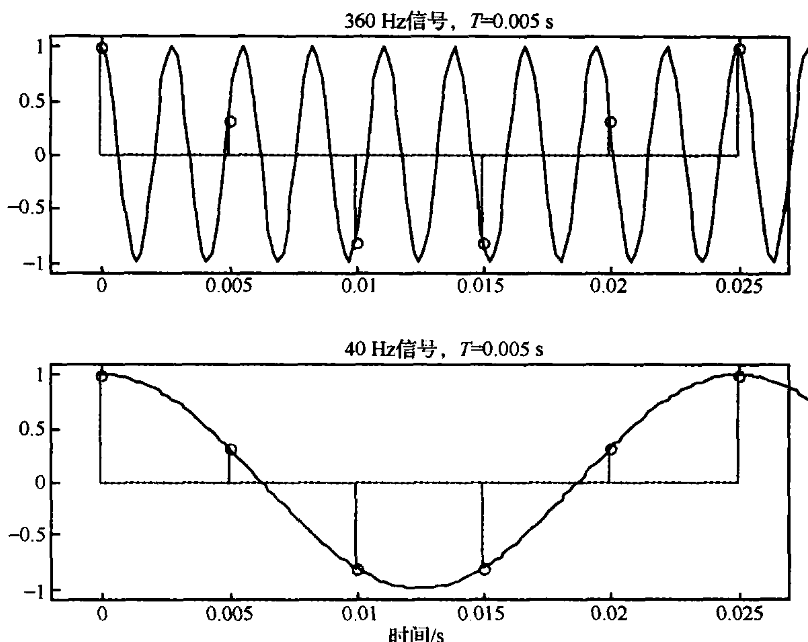


图 10.6 每 5ms 采样一次的 360Hz 正弦波（即采样频率为 200Hz）

注：该采样频率足以采样 40Hz 的正弦波，但不能采样 360Hz 的正弦波。

10.5 生物信号的频域表示

19 世纪初，傅里叶（Joseph Fourier）提出了函数拟合领域最精美的一个数学定理。最初该定理是用于固体热力学传导问题，但随即就被迅速推广。如今，这个傅里叶定理已成为简单函数拟合复杂波形的通用理论，在数学、物理学和工程技术中都得到了广泛的应用。本节将简述傅里叶变换及其各种不同的表达形式，它们对于生物信号的分析和处理具有重要的作用。

10.5.1 周期信号的表示方法——三角傅里叶级数

画家在帆布上画油画时，总是先用颜料在调色板上精心配制出各种不同的颜色，再涂刷到画面上表现不同的色彩。众所周知，红绿蓝（RGB）三原色以适当的比例简单组合之后，就可以显示出彩色色谱上的所有颜色。电视机和计算机的显示通常用 RGB 方式，组合产生颜色，很像画家在画布上所做的工作。实际上，人眼视觉系统的处理过程恰恰与此相反。视网膜将外界的图像和场景的颜色分解成单纯的红、绿、蓝 3 种信号，然后由大脑对这些信号分别进行分析和处理。不过，尽管如此，我们还是能够辨识很多种颜色及其不同的亮暗程度。

这种简单的调色原理其实是傅里叶定理的核心，该定理表明任何复杂波形均可以用简单函数来拟合，并达到所需的任意精度。傅里叶在 1807 年指出，任意一个周期为 T 的周期信号在数学上都可以表示成一组三角函数之和。图 10.7 就显示了方波函数的这种形成过程。理论上，通过同时调节各个正弦波的振幅和频率，就可以实现这种组合。如果振幅和频率选择恰当，三角函数的相加可以组合成任意周期信号。这就像按照精确的比例组合三原色产生任意颜色和任意亮暗的原理一样。RGB 三原色是构成各种精细颜色的基本元素，同理，具有不同频率的正弦波是构成各种复杂信号的基本元素。各种颜色及其所需比例，或者各种正弦波频率及其振幅，都必须精确地调节才能获得理想的结果。例如，用频率和振幅合适的单个正弦波可以实现方波信号的一阶拟合，在其中加入高频正弦波成分（即谐波）可以改善其拟合的效果。如果不断重复这个过程，增加更高频率的正弦波，就可以无限逼近方波信号（见图 10.7）。

随着正弦波数目的增加，拟合质量越来越高。

如下傅里叶级数表示了上述拟合的求和计算：

$$x(t) = a_0 + \sum_{m=1}^{\infty}(a_m \cos m\omega_0 t + b_m \sin m\omega_0 t) \tag{10.3a}$$

式中 $\omega_0 = 2\pi/T$——$x(t)$的基频，单位为弧度/秒，即 rad/s；

a_m 和 b_m——分别是频率为 $\omega_m = m\omega_0$ 的余弦项和正弦项的振幅。

式（10.3a）表明，周期信号 $x(t)$ 可以用无限个正弦波之和精确地重建出来。正弦函数的频率总是 ω_0 的整数倍，ω_0 被称为基波频率，各个正弦函数被称为基波的

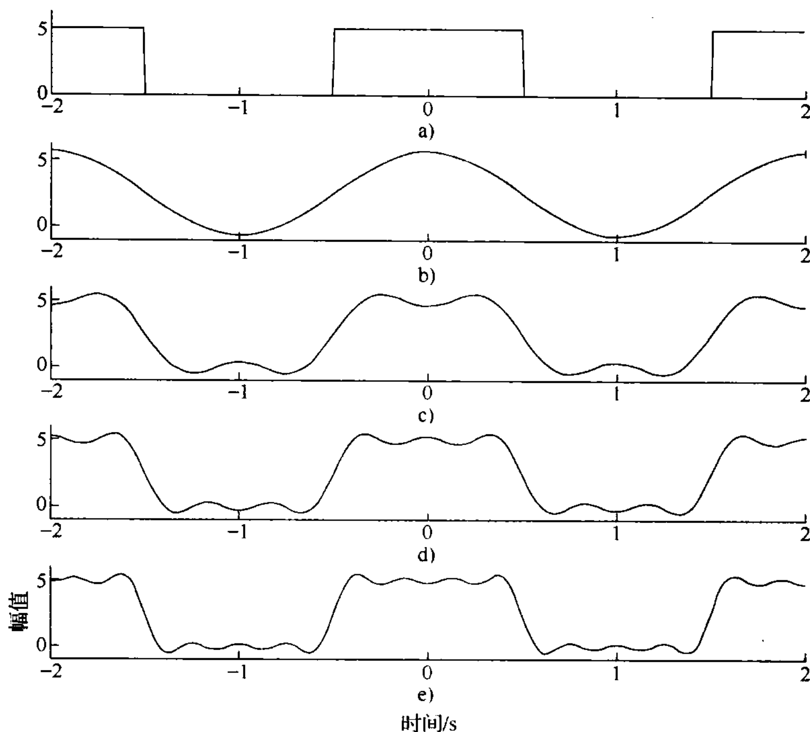


图 10.7　方波信号（图 a）可以用不同数目正弦波的叠加来拟合：1 个正弦波（图 b）、2 个正弦波（图 c）、3 个正弦波（图 d）、4 个正弦波（图 e）。

“谐波”。如果每个正弦或余弦的系数 a_m 和 b_m 已知，求傅里叶级数就可以完全复原信号 $x(t)$。那么，如何确定任意信号的 a_m 和 b_m 呢？

傅里叶级数的系数对应于正弦波的振幅，与每个正弦和余弦的能量相关，可以计算如下：

$$a_0 = \frac{1}{T}\int_T x(t)\mathrm{d}t \tag{10.3b}$$

$$a_m = \frac{2}{T}\int_T x(t)\cos(m\omega_0 t)\mathrm{d}t \tag{10.3c}$$

$$b_m = \frac{2}{T}\int_T x(t)\sin(m\omega_0 t)\mathrm{d}t \tag{10.3d}$$

式中积分区间均为波形的一个周期 T。

例 10.6　请写出图 10.7a 所示方波信号的傅里叶三角级数，并用 MATLAB 程序计算前 10 项的组合结果，画出时间波形图和傅里叶系数图。

解： 首先

$$T=2,\ \omega_0=\frac{2\pi}{T}=\pi$$

为了简化计算，a_m 和 b_m 的积分区间设为方波周期的 $-1\sim1$ 区间，于是

$$a_0 = \frac{1}{T}\int_{-1}^{1} x(t)\mathrm{d}t = \frac{1}{2}\int_{-1/2}^{1/2} 5\mathrm{d}t = \frac{5}{2}$$

$$a_m = \frac{2}{T}\int_{-1}^{1} x(t)\cos(m\omega_0 t)\mathrm{d}t = \int_{-1/2}^{1/2} 5\cos(m\pi t)\mathrm{d}t$$

$$= -5\,\frac{\sin(m\pi t)}{m\pi}\Bigg|_{-1/2}^{1/2} = -5\,\frac{\sin(m\pi/2)}{m\pi/2} = 5\mathrm{sinc}(m\pi/2)$$

$$b_m = \frac{2}{T}\int_{-1}^{1} x(t)\sin(m\omega_0 t)\mathrm{d}t = \int_{-1/2}^{1/2} 5\sin(m\pi t)\mathrm{d}t = -5\,\frac{\cos(m\pi t)}{m\pi}\Bigg|_{-1/2}^{1/2} = 0$$

其中，sinc(x) 的定义是 sinc(x) $=\sin(x)/x$。将 a_0、a_m 和 b_m 的值代入式（10.3a），可得

$$x(t) = \frac{5}{2} + 5\sum_{m=1}^{\infty}\frac{\sin(m\pi/2)}{m\pi/2}\cos(m\pi t)$$

MATLAB 程序如下：

```
%Plotting Fourier Series Approximation
subplot(211)
time=-2:0.01:2; %Time Axis
x = 5/2; %Initializing Signal
for m=1:10
  x = x + 5*sin (m*pi/2)/m/pi *2cos (m*pi*time);
end
plot(time,x,'k') %Plotting and Labels
xlabel('Time (sec)')
ylabel('Amplitude')
set(gca,'Xtick',[-2:2])
set(gca,'Ytick',[0 5])
set(gca,'Box','off')

%Plotting Fourier Magnitudes
subplot(212)
m=1:10;
Am=[5/2 5*sin(m* pi/2)./m/pi*2]; %Fourier Magnitudes
Faxis=(0:10)*.5; %Frequency Axis
plot(Faxis,Am,'k.') %Plotting
axis([0 5 -2 4])
set(gca,'Box','off')
xlabel('Frequency (Hz)')
ylabel('Fourier Amplitudes')
```

该程序运行的输出结果见图 10.8。

注意，图 10.8a 所示的前 10 个谐波的求和结果已经很接近理想方波信号。图 10.8b 显示了前 10 个谐波的傅里叶系数 a_m。要完全重现方波信号快速变化的上升沿和下降沿，就需要无限个谐波的组合。■

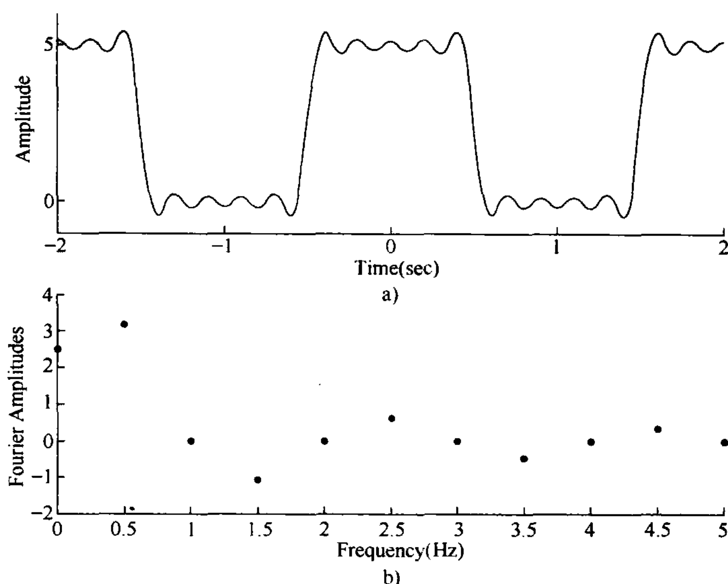


图 10.8　MATLAB 程序计算的

a）方波信号的傅里叶级数前 10 项拟合结果　b）各个谐波频率的傅里叶系数

10.5.2　简化傅里叶级数

傅里叶级数的三角函数形式提供了一种直接拟合和分析各种生物信号的方法，如心脏的重复搏动、人说话时声带所产生的周期性振动等，都可以拟合。傅里叶级数还有其他形式，对于各种不同的应用场合，不同的形式可能更有利于数学处理和信号分析。其中，生物信号拟合和建模中最常用的形式是简化傅里叶级数和指数傅里叶级数。

简化傅里叶级数是标准傅里叶级数的孪生兄弟。由于正弦函数和余弦函数之和可以写成单个包含相位常数的余弦函数，即 $a_m \cos m\omega_0 t + b_m \sin m\omega_0 t = A_m \cos(m\omega_0 t + \phi_m)$，因此，可以导出傅里叶级数的简化形式为

$$x(t) = \frac{A_0}{2} + \sum_{m=1}^{\infty} A_m \cos(m\omega_0 t + \phi_m) \tag{10.4a}$$

各个余弦函数的振幅 A_m 与傅里叶系数之间的关系为

$$A_m = \sqrt{a_m^2 + b_m^2} \tag{10.4b}$$

余弦函数的相位也可以用 a_m 和 b_m 求得，即

$$\phi_m = \arctan\left(\frac{-b_m}{a_m}\right) \tag{10.4c}$$

例 10.7　请将例 10.6 方波函数的标准傅里叶级数转化成简化形式，并编写相应的 MATLAB 程序。

解： 首先，需要计算简化傅里叶级数的振幅 A_m 和相位 ϕ_m，振幅为

$$A_m=\sqrt{a_m^2+b_m^2}=\sqrt{(5\text{sinc}(m\pi/2))^2+(0)^2}=5\frac{|\sin(m\pi/2)|}{\pi m/2}$$

由于

$$|\sin(m\pi/2)|=\begin{cases}1 & m=\text{奇数}\\0 & m=\text{偶数}\end{cases}$$

因此

$$Am=\begin{cases}10/m\pi & m=\text{奇数}\\0 & m=\text{偶数}\end{cases}$$

注意，与标准傅里叶级数的 a_m 和 b_m 不同，对于所有 m，A_m都是正值。相位计算如下：

$$\phi_m=\arctan\left(\frac{-b_m}{a_m}\right)=\arctan\left(\frac{0}{5\text{sinc}(m\pi/2)}\right)=\begin{cases}0 & m=0,1,4,5,8,9\cdots\\\pi & m=2,3,6,7,10,11\cdots\end{cases}$$

于是，整个简化傅里叶级数的表达式为

$$x(t)=\frac{5}{2}+\sum_{m=1}^{\infty}\frac{10}{m\pi}\cos(m\omega_0 t+\phi_m)$$

其中，ϕ_m已在前面定义过。方波的标准傅里叶级数及其简化形式其实很相似。标准形式的系数 a_m 交替取正值和负值，而简化形式的傅里叶系数 A_m的大小与 a_m 相同，但总是正值。标准傅里叶系数的正、负在简化形式中体现为相位交替取 0 和 π。由于 $-\cos(x)=\cos(x+\pi)$，就使得余弦函数的值交替变换正、负号。由此可见，两个方程在数学上其实是相同的，只是三角函数的表示方法不同。

MATLAB 实现程序如下：

```
%Plotting Fourier Series Approximation
time=-2:0.01:2; %Time Axis
x=5/2; %Initializing Signal
m=1:10;
A=(10*sin *(m*pi/2)./m/pi); %Fourier Coefficients
P=angle(A); %Phase Angle
A=abs(A); %Fourier Magnitude
for m=1:10
  x=x+A(m)* cos(m*pi* time+P(m));
end
subplot(211)
plot(time,x,'k') %Plotting and Labels
xlabel('Time (sec)')
ylabel('Amplitude')
set(gca,'Xtick',[-2:2])
set(gca,'Ytick',[0 5])
set(gca,'Box','off')

%Plotting Fourier Magnitudes
subplot(212)
```

```
m=1:10;
A=[5/2 A]; %Fourier Magnitudes
Faxis=(0:10)*.5; %Frequency Axis
plot(Faxis,A,'k.') %Plotting
axis([0 5 -2 4])
set(gca,'Box','off')
xlabel('Frequency (Hz)')
ylabel('Fourier Amplitudes')
```

■

该程序的输出结果与图 10.8a 和 10.8b 所示相同。

10.5.3 指数傅里叶级数

傅里叶级数分析的主要结论是任意周期信号都可以用具有特定振幅和相位的一系列余弦函数之和来拟合。该结论构成了信号分析的基本原理和理论体系。在实际应用中，傅里叶级数是各种准周期信号建模非常有用的工具。

根据 Euler 恒等式 $\cos(\theta)=(e^{j\theta}+e^{-j\theta})/2$ 和 $\sin(\theta)=(e^{j\theta}-e^{-j\theta})/2j$（其中 $j=\sqrt{-1}$），正弦函数和余弦函数都可以用复指数函数来表示。于是，将 Euler 公式应用到简化傅里叶三角级数，任意周期信号就可以用如下复指数函数之和来表示：

$$x(t) = \sum_{m=-\infty}^{+\infty} c_m e^{jm\omega_0 t} \tag{10.5a}$$

该式就是周期信号的指数傅里叶级数。系数 c_m 为复数，它与三角傅里叶级数的系数之间存在如下关系

$$c_m = \frac{a_m - jb_m}{2} = \frac{A_m}{2}e^{j\varphi_m} \tag{10.5b}$$

该式的证明已超出本书的范围。重要的是要知道傅里叶级数的三角形式和指数形式密切相关，比较它们的系数就可以看到这一点。直接由周期信号 $x(t)$ 在一个周期上的积分也可以求得指数系数，即

$$c_m = \frac{1}{T}\int_T x(t)e^{-jm\omega_0 t}dt \tag{10.5c}$$

与三角傅里叶级数一样，只要求取足够数量的复指数函数之和，就可以用傅里叶级数的指数形式以任意精度拟合周期信号。不过，指数傅里叶级数具有一个独特的优点，它只需要计算式 (10.5c) 的一个积分式，而三角形式的傅里叶级数，则需要计算 3 个不同的积分式。

例 10.8　请写出图 10.7a 所示方波信号的指数傅里叶级数，并用 MATLAB 程序计算其前 10 项的组合结果，画出时间波形图和傅里叶系数图。

解： 与例 10.6 相同，傅里叶系数的积分区间取 −1～1。由于这个周期的方波信号只在 −1/2～1/2 有非零值，积分的上下限就可以缩小为 −1/2 和 1/2，即

$$c_m = \frac{1}{T}\int_T x(t)e^{-jm\omega_0 t}dt = \frac{1}{2}\int_{-1/2}^{1/2} 5e^{-jm\pi t}dt$$
$$= \frac{5}{2}\left.\frac{e^{-jm\pi t}}{-jm\pi}\right|_{-1/2}^{1/2} = \frac{5}{2}\frac{e^{jm\pi/2}-e^{-jm\pi/2}}{jm\pi}$$

$$= \frac{5}{2}\frac{\sin(m\pi/2)}{m\pi/2}$$

因此

$$x(t) = \sum_{m=-\infty}^{+\infty} c_m e^{jk\omega_0 t} = \sum_{m=-\infty}^{\infty} \frac{5}{2}\frac{\sin(m\pi/2)}{m\pi/2} e^{jm\pi t}$$

MATLAB 的程序如下：

```
%Plotting Fourier Series Approximation
subplot(211)
time=-2:0.01:2; %Time Axis
x=0; %Initialize Signal
for m=-10:10
   if m==0
   x=x+5/2; %Term for m=0
   else
   x=x+5/2 *sin(m* pi/2)/m/pi*2 **exp(j*m *pi*time);
   end
end
plot(time,x,'k') %Plotting and Labels
xlabel('Time (sec)')
ylabel('Amplitude')
set(gca,'Xtick',[-2:2])
set(gca,'Ytick',[0 5])
set(gca,'Box','off')

%Plotting Fourier Magnitudes
subplot(212)
m=(-10:10)+1E-10;
A=[5/2*sin (m*pi/2)./m/ pi*2]; %Fourier Magnitudes
Faxis=(-10:10)*.5; %Frequency Axis
plot(Faxis,A,'k.') %Plotting
axis([-5 5 -2 4])
set(gca,'Box','off')
xlabel('Frequency (Hz)')
ylabel('Fourier Amplitudes')
```

注意，该 MATLAB 程序的输出结果如图 10.9 所示，这里需要用正、负频率来拟合方波信号。 ■

实际应用中，很多周期信号和准周期信号都只需要用几个谐波分量就可以达到较精确的拟合。图 10.10 和图 10.11 显示了应用傅里叶级数拟合主动脉血压波形所需的谐波分量。其中，图 10.10 是一系列余弦级数谐波的系数图。注意，低频振幅系数的值较大，而高频振幅系数的值较小，也就是其具有的能量较小，对波形重建不起什么作用。图中振幅系数 A_m 的坐标是以 10 为基底的对数，即 lg，因此，小振幅值被放大，也能看得见。图 10.11 显示了几种谐波重建信号。其中，均值加上第一谐波和第二谐波描述了心脏收缩和舒张的基本波形，这两个谐波的振幅很大，因此，对波

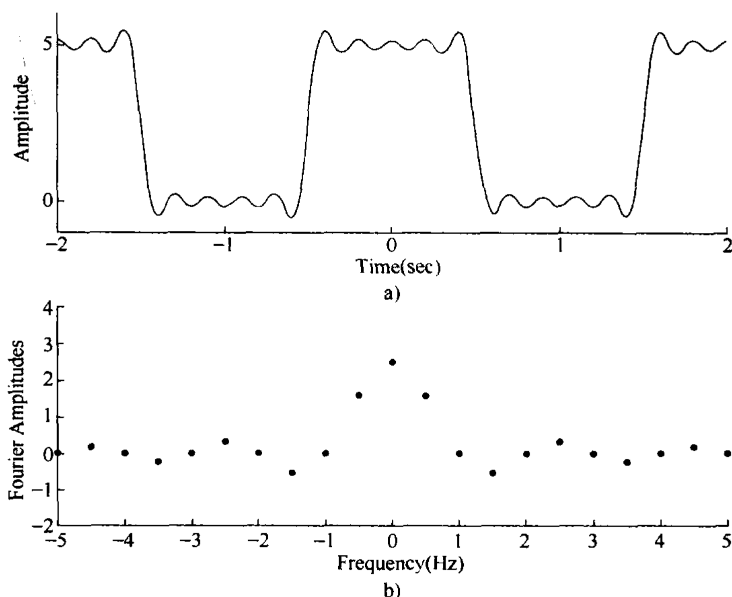


图 10.9　MATLAB 程序计算的结果

a）方波信号的指数傅里叶级数前 10 项拟合结果　b）各个谐波频率的傅里叶系数

注：正、负频率分量合起来才能拟合方波信号。

形重建起了主要作用。其他谐波分量只是添加了波形的一些细节，它们在原始波形中所占能量很小。

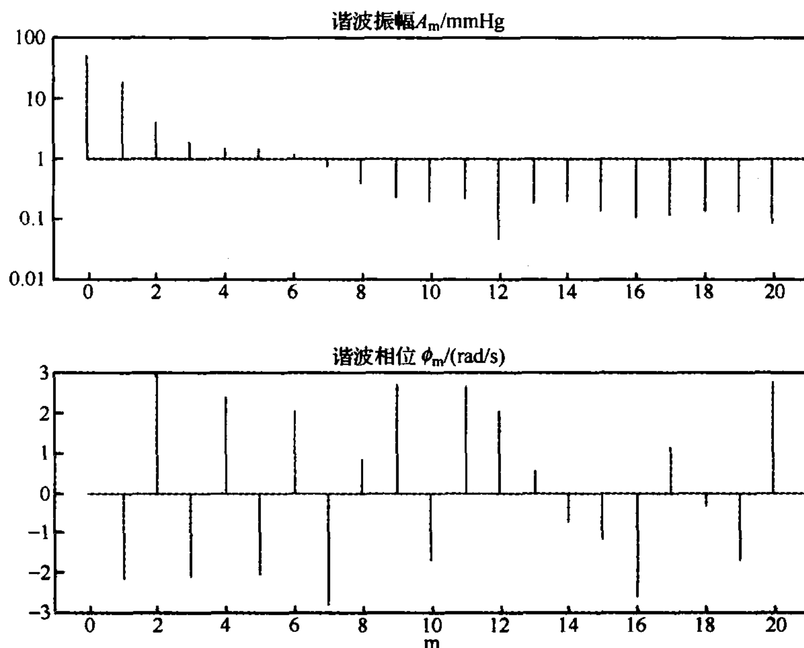


图 10.10　图 10.2 所示主动脉血压波形的谐波系数

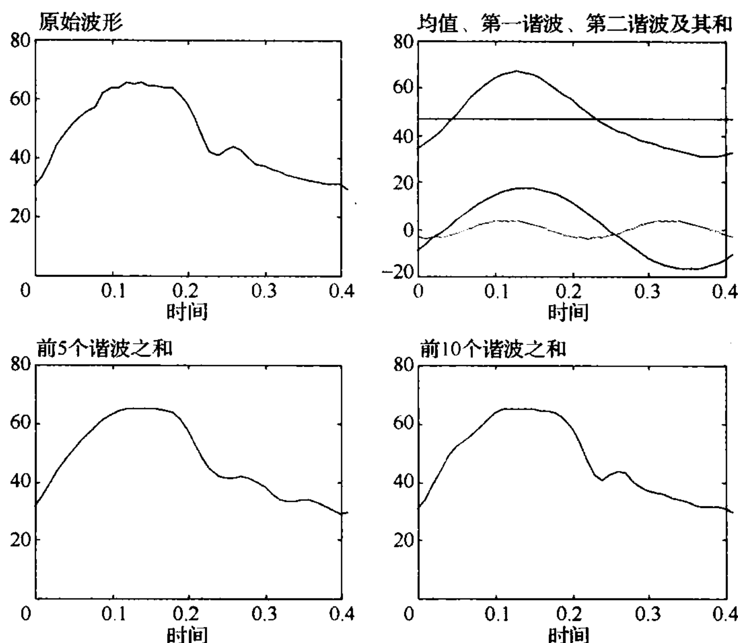


图 10.11 图 10.2 所示主动脉血压波形的谐波重建

10.5.4 傅里叶变换

很多情况下，用余弦函数或正弦函数分量来表示信号具有很多优势。频域这个抽象概念就来自周期信号的傅里叶级数表示方法。信号可以用“时域”的时间函数 $x(t)$ 来表示；也可以用“频域”的谐波频率 $\omega_m = m\omega_0$ 的函数——傅里叶系数 A_m 和相位 ϕ_m 来表示。因此，如果已知信号各个频率分量的傅里叶系数和相位，那么，就完全可以复原周期信号 $x(t)$。

傅里叶级数的一个不足之处就是它只适用于周期信号，而多数生物信号并不是周期信号。实际上，一般生物信号都是时间的连续函数，且决不会重复。幸运的是，用余弦、正弦或者复指数来拟合周期信号的傅里叶级数的概念还可以扩展到非周期信号。傅里叶积分，也称为傅里叶变换（Fourier Transform，FT），就是用于将连续的非周期信号分解成不同频率分量的组合，即

$$X(\omega) = \int_{-\infty}^{\infty} x(t)\mathrm{e}^{-j\omega t}\mathrm{d}t \tag{10.6}$$

这种方法很像傅里叶级数将周期信号分解成其相应的三角函数分量。其中，$X(\omega)$ 是连续频率 ω 的复数函数，与复数傅里叶级数的系数 c_m 类似。该关系式的严格推导和证明超出了本书的范围，但要注意，傅里叶积分可以由指数傅里叶级数直接推导，只要将其周期 T 趋于无穷大就可以了。当 $T\rightarrow\infty$ 时，三角函数级数的系数 c_m

就趋于 $X(\omega)$。理论上，周期无穷大的函数就可以看作非周期函数，因为你永远看不到其重复。一般数字信号处理的教科书中都可以查到常见信号的傅里叶变换计算公式。

与傅里叶级数一样，这里也需要将频域信号 $X(\omega)$ 转换为其对应的时域信号。用傅里叶反变换（Inverse Fourier Transform，IFT）就可以从傅里叶变换中恢复出时域信号 $x(t)$，即

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{j\omega t}d\omega \tag{10.7}$$

这两种信号的表达可以互换，也就是时域信号 $x(t)$ 与频域信号 $X(\omega)$ 之间总是可以变来变去。因此，频域信号其实包含了信号的所有信息，使人们可以从另一个不同角度来分析和处理生物信号。

例 10.9　请计算如下方波脉冲信号的傅里叶变换

$$x(t)=\begin{cases}1, & |t|\leqslant a\\ 0, & |t|>a\end{cases}$$

解：应用式（10.6），可得

$$X(\omega) = \int_{-a}^{a} e^{-j\omega t}dt = \left.\frac{e^{-j\omega t}}{-j\omega}\right|_{-a}^{a} = \frac{2\sin\omega a}{\omega} = 2a\text{sinc}(\omega a)$$ ■

就像用傅里叶级数表示信号一样，幅值和相位是傅里叶变换的重要特性。如前所述，$X(\omega)$ 是一个复函数，由实部 $\text{Re}\{X(\omega)\}$ 和虚部 $\text{Im}\{X(\omega)\}$ 构成，可以表示为

$$X(\omega)=\text{Re}\{X(\omega)\}+j\text{Im}\{X(\omega)\} \tag{10.8}$$

与傅里叶级数相似，式（10.8）的取值决定了从其傅里叶变换重建原信号 $x(t)$ 所需要的每个复指数函数（即等价于余弦函数）的幅值，即

$$|X(\omega)|=\sqrt{\text{Re}\{X(\omega)\}^2+\text{Im}\{X(\omega)\}^2} \tag{10.9}$$

并且，相位决定了每个余弦信号相对于零时间参考点的时间偏移，其计算如下：

$$\theta(\omega)=\arctan\left(\frac{\text{Im}\{X(\omega)\}}{\text{Re}\{X(\omega)\}}\right) \tag{10.10}$$

注意，这里的幅值和相位计算公式与简化傅里叶三角级数的振幅和相位的计算公式（式 10.4a、b、c）很相似，其中，傅里叶变换的幅值 $|X(\omega)|$ 与 A_m 类似，而 a_m 和 b_m 则分别与 $\text{Re}\{X(\omega)\}$ 和 $\text{Im}\{X(\omega)\}$ 类似。

例 10.10　请计算如下傅里叶变换信号的幅值和相位

$$X(\omega)=\frac{1}{1+j\omega}$$

解：先将信号转化为与式（10.8）相似的形式，即

$$X(\omega)=\frac{1}{1+j\omega}\,\frac{1-j\omega}{1-j\omega}=\frac{1-j\omega}{1+\omega^2}=\frac{1}{1+\omega^2}-j\,\frac{\omega}{1+\omega^2}$$

于是

$$\mathrm{Re}\{X(\omega)\}=\frac{1}{1+\omega^2},\mathrm{Im}\{X(\omega)\}=-\frac{\omega}{1+\omega^2}$$

应用式（10.9）和式（10.10），其幅值即为

$$|X(\omega)|=\frac{1}{1+\omega^2}$$

相位为

$$\theta(\omega)=\arctan(-\omega)=-\arctan(\omega)$$ ■

10.5.5 傅里叶变换的性质

实际应用中，复杂信号的傅里叶变换（FT）是一项繁重且费时的工作，需要借助某些方法简化计算过程。利用 FT 所具有的几个特性就可以简化时域信号与频域信号之间的转换，简介如下。

设 $x_1(t)$ 和 $x_2(t)$ 是两个时域信号，它们的 FT 分别为 $X_1(\omega)=F\{x_1(t)\}$ 和 $X_2(\omega)=F\{x_2(t)\}$。

1. 线性性质

傅里叶变换是一个线性算子，因此，对于任意常数 a_1 和 a_2，有

$$F\{a_1x_1(t)+a_2x_2(t)\}=a_1X_2(\omega)+a_2X_2(\omega) \tag{10.11}$$

该式说明，线性系统的比例运算和求和运算性质同样适用于傅里叶变换。

2. 时移（即延时）性质

如果 $x_1(t-t_0)$ 是具有某个时移 t_0 的时域信号，则其傅里叶变换为

$$F\{x_1(t-t_0)\}=e^{-j\omega t_0}X_1(\omega) \tag{10.12}$$

也就是，信号在时间上的位移对应于其傅里叶变换乘以一个相位因子 $e^{-j\omega t_0}$。

3. 频移性质

如果 $X_1(\omega-\omega_0)$ 是在频域上具有位移 ω_0 的某信号的傅里叶变换，那么，其傅里叶反变换（IFT）为

$$F^{-1}\{X_1(\omega-\omega_0)\}=e^{j\omega_0 t}x(t) \tag{10.13}$$

4. 卷积定理

两个时域信号的卷积定义为

$$x(t)=\int_{-\infty}^{\infty}x_1(\tau)x_2(t-\tau)\mathrm{d}\tau=x_1(t)*x_2(t) \tag{10.14}$$

其频域的等价运算为

$$X(\omega)=F\{x(t)\}=F\{x_1(t)*x_2(t)\}=X_1(\omega)X_2(\omega) \tag{10.15}$$

由此可见，在时域中难以计算的卷积运算在频域中就是乘积运算。

下面再来看两个频域信号 $X_1(\omega)$ 和 $X_2(\omega)$ 在频域中的卷积积分，即

$$X(\omega)=\int_{-\infty}^{\infty}X_1(v)X_2(\omega-v)\mathrm{d}v=X_1(\omega)*X_2(\omega) \tag{10.16}$$

$X(\omega)$ 的 IFT 为

$$x(t)=F^{-1}\{X(\omega)\}=F^{-1}\{X_1(\omega)*X_2(\omega)\}=2\pi x_1(t)x_2(t) \tag{10.17}$$

因此，频域中两个信号的卷积对应于时域中这两个信号的乘积乘以系数 2π。后面我们将会看到，卷积是生物信号滤波处理的重要方法。

例 10.11　请计算 $3\sin(25t)+4\cos(50t)$ 的傅里叶变换，只要用符号表达式表示即可，不用计算其结果。

解：　$F\{3\sin(25t)+4\cos(50t)\}=3F\{\sin(25t)\}+4F\{\cos(50t)\}$ ■

10.5.6 离散傅里叶变换

在数字信号应用中，连续的生物信号首先要用模/数转换器采样，由计算机读入，然后再进行分析和处理。上述傅里叶变换只适用于连续时间信号，如果要用它分析离散信号的频域特性，那么，就先要改变傅里叶变换公式，使其能够用于离散的采样数据。

离散傅里叶变换（Discrete Fourier Transform，DFT）的计算公式为

$$X(m)=\sum_{k=0}^{N-1}x(k)\mathrm{e}^{-\mathrm{j}\frac{2\pi mk}{N}};m=0,1,\cdots,N/2 \tag{10.18}$$

它是频域上描述和分析离散信号必不可少的工具。DFT 其实是傅里叶变换的数字化形式。下标 m 表示数字频率的序号，$x(k)$ 是 $x(t)$ 的采样逼近，k 是离散时间变量，N 是一个偶数，是 $x(k)$ 的采样点数，$X(m)$ 就是 $x(k)$ 的 DFT。

同理，离散傅里叶反变换（Inverse Discrete Fourier Transform，IDFT）是傅里叶反变换的离散形式，其计算公式为

$$x(k)=\frac{1}{N}\sum_{m=0}^{N-1}X(m)\mathrm{e}^{\mathrm{j}\frac{2\pi mk}{N}};k=0,1,\cdots,N-1 \tag{10.19}$$

与 FT 和 IFT 之间的关系一样，DFT 和 IDFT 是离散域的傅里叶变换对。DFT 可以将一系列离散时间点上的采样数据转换成其频域表示形式。相反，IDFT 则实现 DFT 的逆过程，可以从信号的频域形式 $X(m)$ 重建采样信号 $x(k)$。这两个公式可以互换，每个都包含了离散信号的所有信息。

例 10.12　请计算信号 $x(k)=0.25^k$ 的离散傅里叶变换，其中 $k=0\sim15$。

解：　$$X(m)=\sum_{k=0}^{N-1}x(k)\mathrm{e}^{-\mathrm{j}\frac{2\pi mk}{N}}=\sum_{k=0}^{15}0.25^k\mathrm{e}^{-\mathrm{j}\frac{2\pi mk}{16}}=\sum_{k=0}^{15}(0.25\mathrm{e}^{-\mathrm{j}\frac{2\pi mk}{N}})^k=\sum_{k=0}^{15}a^k$$

注意，该式是一个几何级数的求和，其中 $a=0.25\mathrm{e}^{-\mathrm{j}\frac{2\pi m}{N}}$。于是有

$$\sum_{k=M}^{N}a^k=\frac{a^{N+1}-a^M}{a-1}$$

可得

$$X(m)=\frac{a^{16}-a^0}{a-1}=\frac{0.25^{16}\mathrm{e}^{-\mathrm{j}\frac{32m\pi}{N}}-1}{0.25\mathrm{e}^{-\mathrm{j}\frac{2m\pi}{N}}-1}$$ ■

快速傅里叶变换（Fast Fourier Transform，FFT）是求取 DFT 的一种有效的计算机算法。FFT 算法的输出结果与 DFT 算法相同，只是 FFT 的运算速度比 DFT 快得多。两种算法的运算时间分别相当于 $N\cdot\log_2(N)$ 数量级和 N^2 数量级，因此，两

者的计算时间之比为

$$\frac{\text{DFT 的计算时间}}{\text{FFT 的计算时间}}=\frac{N^2}{N\cdot\log_2 N}=\frac{N}{\log_2 N} \tag{10.20}$$

不过，只有当采样数 N 为 2 的整数次幂时，FFT 才有效。如果 $N=1024$ 个采样点，则 FFT 算法比直接用 DFT 快大约 10 倍（$1024/\log_2(1024)=10$）。如果 N 不是 2 的整数次幂，则通常采用 DFT 算法。

图 10.12a 所示是频率为 100 Hz 的正弦波信号，图 10.12b 是用 FFT 算法求得的该正弦波的 DFT。注意，FFT 的峰值就在正弦波频率 100 Hz 处。图 10.13a 所示是叠加了随机噪声的 100 Hz 正弦波，正弦波频率在时域中似乎看不出来。但是，将信号转换到频域之后，100 Hz 处的频率分量就显示出来了，其 FFT 的计算结果中存在一个大峰值（见图 10.13b）。

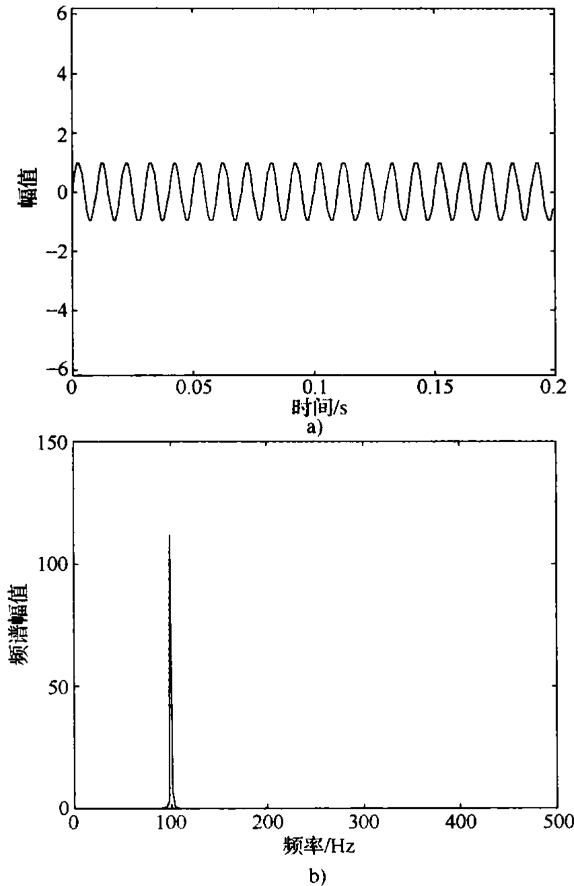


图 10.12 100 Hz 正弦波（图 a）和该正弦波的快速傅里叶变换（图 b）

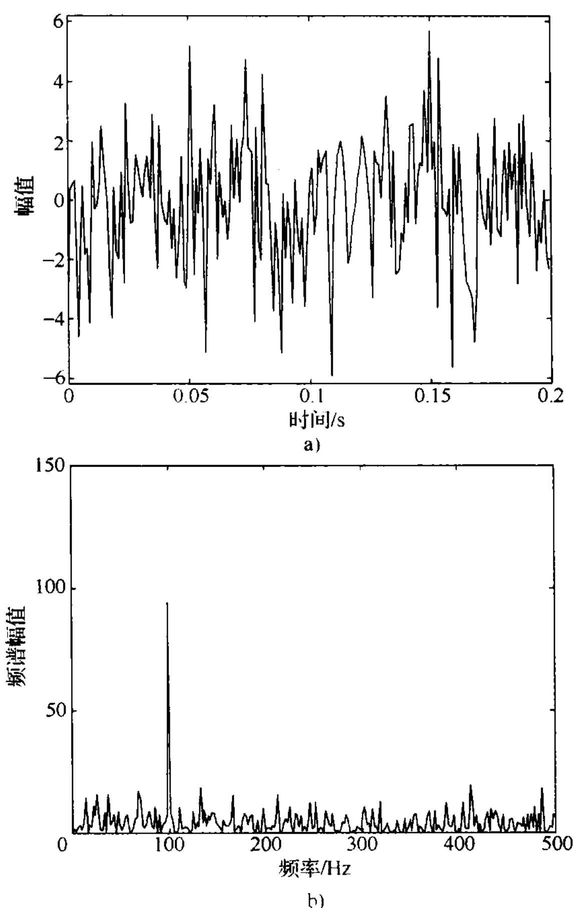


图 10.13 含随机噪声的 100 Hz 正弦波（图 a）和该信号的快速傅里叶变换（图 b）

例 10.13 请编写 MATLAB 程序，计算信号 $x(n)=\sin(\pi/4n)+2\cos(\pi/3n)$ 的离散傅里叶变换的幅值，并作图。

解：MATLAB 程序如下：

```
n = 1:1024; %Discrete Time Axis
x = sin (pi/4*n) + 2*cos (pi/3*n); %Generating the signal
X = fft(x, 1024*16)/1024; %Computing 16k point Fast Fourier Transform
Freq = (1:1024*16)/(1024*16)*2*pi; %Normalizing Frequencies between 0 - 2*pi
plot(Freq,abs(X),'k') %Plotting
axis([0.7 1.15 0 1.2])
xlabel('Frequency (rad/s)')
ylabel('Fourier Magnitude')
```

该程序的输出结果如图 10.14 所示。

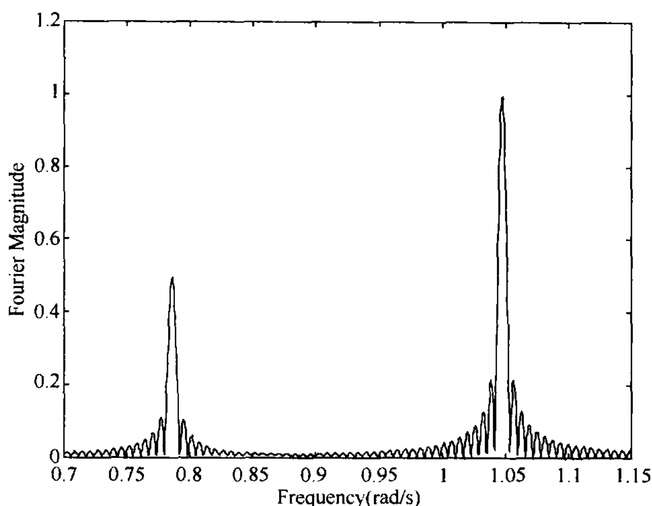


图 10.14 两个正弦波之和的快速傅里叶变换的幅值，其主要能量集中在 π/3 和 π/4rad/s 两个频率处。

■

10.5.7 Z 变换

Z 变换是另一种离散信号分析方法。该变换其实是由 DFT 衍变而来，它将离散数据序列转换到 Z 域。在多数应用情况下，Z 变换比 DFT 容易操作，因为它不用复数。Z 变换对于数字信号的作用相当于拉普拉斯变换对于连续信号的作用。

如果用 x_k 表示离散序列 $x(k)$，那么，该离散序列的（单边）Z 变换计算公式就是

$$X(z) = \sum_{k=0}^{\infty} x_k z^{-k} = x_0 + x_1 z^{-1} + x_2 z^{-2} + \cdots \tag{10.21}$$

注意，Z 变换可以直接从 DFT 得到。令式（10.18）的 $N \to \infty$ 并且使 $z = e^{-j\frac{2\pi m}{N}}$ 即可。实际应用中，采样所得的生物信号都表示为具有 N 个数据点的序列，因此，Z 变换只要计算 $k=0$，…，$N-1$ 就可以了。有关数字信号处理的教科书中都可以找到常见 Z 变换及其反变换的计算公式。

连续信号采样得到离散数据序列之后，其 Z 变换就很容易求得。采样信号的数据序列可以表示为

$$x = [x(0), x(T), x(2T), \cdots, x(kT)] \tag{10.22}$$

应用式（10.21）就可以求得其 Z 变换，即

$$X(z) = x(0) + x(T)z^{-1} + x(2T)z^{-2} + \cdots + x(kT)z^{-k} \tag{10.23}$$

式中 T——采样周期，也就是采样时间间隔。

采样信号是一个数据序列，其中每个采样点与它相邻采样点之间的间隔就是一个采样周期。在 Z 变换中，乘数 $x(kT)$ 是采样数据点的幅值。z^{-k} 项具有直观的含

义。幂 k 对应于采样起始时刻零时间点之后已经过去的采样周期数；因此，z^{-k}可以看作“时移算子”，它将采样点延迟 k 个采样周期，也就是 kT。例如，变量 z^{-1}表示与零时刻有一个周期的延时。式（10.23）中，$z(0)$ 是 $t=0$ 时刻采样点的值，$x(T)$ 是第一个采样周期之后的采样点的值。因此，Z 变换其实是描述模拟信号采样过程的一种重要方法。

例 10.14　离散的单位冲激函数可以用序列 $x=[1, 0, 0, 0, \cdots, 0]$ 来表示。请计算该序列的 Z 变换。

解：　$X(z)=1+0z^{-1}+0z^{-2}+\cdots+0z^{-k}=1+0+0+\cdots+0=1$ ■

例 10.15　神经纤维中记录的电信号用 A/D 转换器转换成数字信号。假设该信号前 5 个采样点的值为 [−60.0，−49.0，−36.0，−23.0，−14.0] mV。请计算该数据序列的 Z 变换。采样开始之后经过几个采样周期才是−23.0mV 这个数据点？

解：　$Y(z)=-60.0-49.0z^{-1}-36.0z^{-2}-23.0z^{-3}-14.0z^{-4}$

−23.0 mV 数据点的 z 变量的负指数值是 3，因此该数据点是在采样开始之后 3 个采样周期后记录的点。■

10.5.8　Z 变换的性质

Z 变换具有许多上述傅里叶变换中已经介绍过的性质，这些性质可用于简化复杂信号的 Z 变换计算，介绍如下。注意，这些性质与式（10.11）、式（10.12）和式（10.14）所述性质很相似。

假设 $x_1(k)$ 和 $x_2(k)$ 是两个数字信号，$X_1(z)$ 和 $X_2(z)$ 是它们各自的 Z 变换。

1. 线性性质

Z 变换是线性算子，因此，对于任意常数 a_1 和 a_2 有

$$
\begin{aligned}
Z\{a_1x_1(k)+a_2x_2(k)\} &= \sum_{k=0}^{\infty}[a_1x_1(k)+a_2x_2(k)]z^{-k} \\
&= a_1X_1(z)+a_2X_2(z)
\end{aligned} \tag{10.24}
$$

2. 延时性质

假设 $x_1(k-n)$ 是原始信号延迟了 n 个采样点的信号，则该延时信号的 Z 变换是

$$
Z\{x_1(k-n)\} = \sum_{k=0}^{\infty}x_1(k-n)z^{-k} = \sum_{k=0}^{\infty}x_1(k)z^{-(k+n)} = z^{-n}X_1(z) \tag{10.25}
$$

如前所述，注意，算子 z^{-n}表示 n 个采样点的时移，也就是严格的 nT 秒时间。

3. 卷积

设 $x(k)$ 是 $x_1(k)$ 与 $x_2(k)$ 的离散卷积，即

$$x(k)=x_1(k)*x_2(k)$$

则 $x(k)$ 的 Z 变换 $X(z)$ 为

$$X(z)=Z\{x(k)\}=X_1(z)X_2(z) \tag{10.26}$$

与傅里叶变换一样，该结果说明，两个数据序列之间的卷积在 z 域中是简单的乘积。

10.6 线性系统

系统可以是一种处理方法、一台机器或者一个装置。它接收输入信号，经过一定的处理之后产生输出信号。其输出与输入相关，却又有明显的区别。图 10.15 所示是系统的各种框图表示法。

生物体的各种器官和结构经常用不同的系统来模拟。例如，心脏就是一个大型系统，它吸收缺氧的静脉血（也就是输入），并将血液泵入肺部；然后将肺部产生的含氧丰富的血液收回心脏，再经过主动脉输出。大脑神经元也可以简化为一个微型系统，它接收来自其他神经元的电脉冲输入信号，将这些输入信号求和，然后产生动作电位作为输出响应信号。线性系统是一类特殊的系统，它所具有的性质使得系统的分析变得很容易。

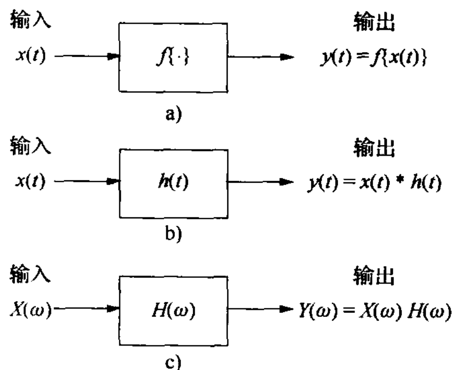


图 10.15 系统的框图表示法：（图 a）输入信号 $x(t)$ 经过系统变换函数 $f\{\cdot\}$的处理后产生输出 $y(t)$；线性系统的时域表示（图 b），这种系统的输出等于输入信号与系统冲激响应的卷积；线性系统的频域表示（图 c），输出等于输入与系统传递函数的乘积

10.6.1 线性系统的性质

尽管生物系统本质上并不是线性系统，但是，通常可以用线性系统模型近似，因为这样可以使系统的分析和处理比较容易进行。

所有线性系统都遵循叠加原理和比例特性。叠加原理表明，两个独立输入之和所产生的输出等于每个输入各自的输出之和。比例特性则表明，输入大小的变化会引起输出的同比例变化。数学上，如果已知两个独立输入信号各自的输出，即

输入　　输出

$$x_1(t) \rightarrow y_1(t)$$

$$x_2(t) \rightarrow y_2(t)$$

那么，就可以方便地计算出这两个输入的任意组合信号所产生的输出，即，输入信号的线性叠加和比例变化所产生的输出信号也是线性叠加和同比例变化：

输入　　　　　　　　输出

$$k_1 \cdot x_1(t) + k_2 \cdot x_2(t) \rightarrow k_1 \cdot y_1(t) + k_2 \cdot y_2(t) \quad (10.27)$$

其中，k_1和 k_2是任意的幅值比例系数。这些系数使输入的幅值放大（$k>1$）或缩小（$k<1$），并使得净输出也产生相应的变化，以相同的比例系数缩放。

例 10.16 以下是一个线性系统的输入和输出对应信号：

输入　　　　　　　　输出

$$x_1(t) = \cos(t) \quad \rightarrow y_1(t) = \cos(t + \pi/2)$$

$$x_2(t)=\cos(t)+\sin(2t) \to y_2(t)=\cos(t+\pi/2)+5\sin(2t)$$

$$x_3(t)=\cos(3t) \qquad \to y_3(t)=2\cos(3t)$$

如果输入为 $x(t)=3\sin(2t)+1/2\cos(3t)$，请推导其输出。

解：首先将输入表示为 x_1、x_2和 x_3的叠加，即：

$$x(t)=3(x_2(t)-x_1(t))+1/2x_3(t)=3x_2(t)-3x_1(t)+1/2x_3(t)$$

然后，应用叠加原理和比例特性，可得其输出为

$$y(t)=3y_2(t)-3y_1(t)+1/2y_3(t)=3[\cos(t+\pi/2)+5\sin(2t)]-3\cos(t+\pi/2)$$
$$+1/2(2\cos(3t))=15\sin(2t)+\cos(3t)$$ ■

例 10.17　假设有如下的系统表达式

$$y(t)=f\{x(t)\}=A\cdot x(t)+B$$

请判断该系统是不是线性系统。

解：为了回答这个问题，假设有两个独立的输入信号 $x_1(t)$ 和 $x_2(t)$，它们各自的输出分别为 $y_1(t)$ 和 $y_2(t)$。将这两个输入信号的叠加 $x(t)=k_1\cdot x_1(t)+k_2\cdot x_2(t)$ 输入系统，如果系统是线性的，则其输出应该是

$$y_{Lin}(t)=k_1\cdot y_1(t)+k_2\cdot y_2(t)=k_1\cdot(A\cdot x_1(t)+B)+k_2\cdot(A\cdot x_2(t)+B)$$
$$=A(k_1\cdot x_1(t)+k_2\cdot x_2(t))+(k_1+k_2)B$$

但是，本题系统的实际输出是

$$y(t)=f\{x(t)\}=f\{k_1\cdot x_1(t)+k_2\cdot x_2(t)\}=A\cdot(k_1\cdot x_1(t)+k_2\cdot x_2(t))+B$$

仔细比较预期的线性系统输出 $y_{Lin}(t)$ 与实际输出 $y(t)$，可见，$y(t)\neq y_{Lin}(t)$，因此该系统不是线性的。 ■

叠加原理对于周期信号具有特殊的意义。由于任意周期信号都可以用傅里叶级数表示为余弦或者复指数函数之和，因此，满足叠加原理的系统的输出也一定是余弦函数或者复指数函数之和。于是，如果线性系统的输入为周期信号，那么，其输出也是具有相同谐波频率的周期信号，即线性系统的周期性输入信号 $x(t)$ 与其输出 $y(t)$ 的关系为

$$\underset{\text{输入}}{x(t)=\frac{A_0}{2}+\sum_{m=-\infty}^{+\infty}A_m\cos(m\omega_0 t+\phi_m)}\Rightarrow y(t)$$

$$\underset{\text{输出}}{=\frac{B_0}{2}+\sum_{m=-\infty}^{+\infty}B_m\cos(m\omega_0 t+\theta_m)} \tag{10.28}$$

输入和输出都包含同样频率 $m\omega_0$ 的余弦函数，两者的表达式也相似。指数傅里叶级数也是如此，即

$$\underset{\text{输入}}{x(t)=\sum_{m=-\infty}^{+\infty}c_m e^{jm\omega_0 t}}\Rightarrow \underset{\text{输出}}{y(t)=\sum_{m=-\infty}^{+\infty}b_m e^{jm\omega_0 t}} \tag{10.29}$$

该式中的输入、输出系数 c_m 和 b_m 显然与式（10.28）的 A_m 和 B_m 相关（参见式

(10.5b))。

由式(10.28)和式(10.29)可见，线性系统周期信号的输入与输出之间存在两个区别。第一，每个余弦函数的振幅系数不同，输入是A_m而输出是B_m。这些系数完全由线性系统的性质决定。第二，输入分量的相角ϕ_m与输出分量的相角θ_m不同，这意味着输入分量与输出分量之间存在时移。与振幅一样，输入与输出之间的相位差也是由线性系统决定的。因此，如果已知输入转化为输出时输入信号振幅变化和相位变化的数学关系式，那么，就可以完整地描述线性系统。这个数学关系就是系统传递函数H_m，它完全可以描述线性系统通过振幅和相位的变化将输入转化为特定输出的过程，这种转化只要用两组独立的参数——幅值和相位就可以表示。

H_m的幅值就是第m个分量的输出与输入之比，即

$$|H_m|=\frac{B_m}{A_m} \tag{10.30}$$

注意，如果第m个分量的输入振幅已知，那么，将传递函数的幅值乘以输入的傅里叶系数（即振幅）就可以得到输出的傅里叶系数，即$B_m=|H_m|\cdot A_m$。传递函数的相位角则描述了第m个频率分量的输入与输出之间的相位关系，即

$$\angle H_m=\theta_m-\phi_m \tag{10.31}$$

如果输入相位已知，就可以求得输出相位，即$\theta_m=\angle H_m+\phi_m$。式(10.30)和式(10.31)是描述线性系统的两个必不可少的关键参数，如果这两个参数已知，就可以计算任意输入的系统输出响应了。

10.6.2 线性系统的时域表示

线性系统输入与输出之间的关系可以用其时域特性来描述，冲激响应函数$h(t)$就可以完整地描述线性系统的特性（见图10.15b）。下面我们就会看到，线性系统的冲激响应与上述周期信号的传递函数直接相关。

如果系统的$h(t)$已知，就可以利用如下卷积积分直接计算任意输入信号$x(t)$的输出响应$y(t)$：

$$y(t)=h(t)*x(t)=\int_{-\infty}^{\infty}h(\tau)x(t-\tau)\mathrm{d}\tau \tag{10.32}$$

式中，符号“*”表示输入与系统冲激响应的卷积，积分变量是τ。对于离散信号，离散线性系统的输出用卷积和来计算，即

$$y(k)=h(k)*x(k)=\sum_{m=-\infty}^{\infty}h(m)x(k-m) \tag{10.33}$$

式中　$h(m)$——离散系统的冲激响应。

有关卷积积分的详细内容在很多信号处理教科书中都有介绍，本书不做深入的介绍。

例 10.18　向细胞内施加单位阶跃电流$i(t)=u(t)$会引起细胞膜电压$v(t)$的变化。假设细胞膜为线性系统，其冲激响应$h(t)=A\mathrm{e}^{-t/\tau}$，其中$A$是常数，单位是V/s/A；$\tau$是细胞膜的时间常数，单位是s。请计算细胞膜的电压输出。

解：根据式（10.32），输入 $i(t)$ 与输出 $v(t)$ 之间的关系是如下卷积积分：

$$v(t)=h(t)*i(t)=\int_{-\infty}^{\infty}h(\zeta)i(t-\zeta)\mathrm{d}\zeta=\int_{-\infty}^{\infty}A\mathrm{e}^{-\zeta/\tau}u(\zeta)u(t-\zeta)\mathrm{d}\zeta$$

式中积分变量用ζ表示，以便与细胞时间常数 τ 区别开来。积分号内的单位阶跃函数取值为 1 或 0，取 0 时对积分没有贡献。当$\zeta>1$ 时，$u(\zeta)=1$；当 $t-\zeta>0$ 时，$u(t-\zeta)=1$。结合这两个不等式，可知积分区间为$0<\zeta<t$，在此区间内单位阶跃函数可以用 1 来代替，因此

$$v(t)=\int_0^t A\mathrm{e}^{-\zeta/\tau}\mathrm{d}\zeta=\frac{A}{\tau}(1-\mathrm{e}^{-t/\tau})$$ ■

10.6.3　线性系统的频域表示

前面已经提到过线性系统周期性输入信号频域分析的特殊情况，即线性系统周期性输入信号的输出响应完全可以用频域的系统传递函数来描述。同时，线性系统的输出也可以用时域的冲激响应卷积积分来表示（见图 10.15b）。线性系统频域和时域上的这两种输入输出关系的描述方法相互之间存在联系。显然，对于非周期信号，传递函数是冲激响应的傅里叶变换，即

$$\begin{matrix}\text{时域} & & \text{频域}\\ h(t) & \Leftrightarrow & H(\omega)\end{matrix}\tag{10.34}$$

式中　$h(t)$——系统的冲激响应；

$H(\omega)$——系统的传递函数。

由于冲激响应是线性系统的完整描述，并且傅里叶变换是可逆的，总是可以在 $h(t)$ 与 $H(\omega)$ 之间互换，因此，传递函数包含了描述系统所需的全部信息。传递函数的优点是容易使用。在很多应用中，卷积积分很难计算。而线性系统在频域的输出可以用系统的输入与其传递函数的乘积来计算，即 $Y(\omega)=X(\omega)H(\omega)$，不用计算卷积积分。这个结果很像式（10.30）和式（10.31）的傅里叶级数。尤其是，傅里叶变换的卷积性质表明，时域的卷积对应于频域的乘积，即

$$\begin{matrix}\text{时域} & & \text{频域}\\ y(t)=x(t)*h(t) & \Leftrightarrow & Y(\omega)=X(\omega)H(\omega)\end{matrix}\tag{10.35}$$

因此，线性系统的输出 $Y(\omega)$ 在频域中就等于 $X(\omega)$ 与 $H(\omega)$ 的乘积（见图 10.15c）。注意，这个结果其实就是式（10.14）和式（10.15）表示的卷积定理在线性系统中的应用（式（10.32））。在很多情况下，这比直接计算时域的卷积要简单得多。下面的例 10.19 将证明卷积定理。

例 10.19　请证明傅里叶变换的卷积特性。

解：　$$Y(\omega)=FT\{y(t)\}=\int y(t)\mathrm{e}^{-\mathrm{j}\omega t}\mathrm{d}t=\iint x(\tau)h(t-\tau)\mathrm{d}\tau\mathrm{e}^{-\mathrm{j}\omega t}\mathrm{d}t$$

进行变量代换，即 $u=t-\tau$，$\mathrm{d}u=\mathrm{d}t$，可得

$$Y(\omega)=\iint x(\tau)h(u)\mathrm{d}\tau\mathrm{e}^{-\mathrm{j}\omega(u+\tau)}\mathrm{d}u=\int x(\tau)\mathrm{e}^{-\mathrm{j}\omega t}\mathrm{d}\tau\cdot\int h(u)\mathrm{e}^{-\mathrm{j}\omega u}\mathrm{d}u=X(\omega)H(\omega)$$ ■

例 10.20 与例 10.18 相同，向细胞内施加刺激电流，请计算频域中的细胞膜输出电压。

解： 阶跃电流输入函数的傅里叶变换为

$$I(\omega)=\int i(t)\mathrm{e}^{-\mathrm{j}\omega t}\mathrm{d}t=\int_0^{\infty}1\cdot \mathrm{e}^{-\mathrm{j}\omega t}\mathrm{d}t=\left.\frac{\mathrm{e}^{-\mathrm{j}\omega t}}{-\mathrm{j}\omega}\right|_0^{\infty}=\frac{1}{\mathrm{j}\omega}$$

传递函数等于冲激响应的傅里叶变换，即

$$H(\omega)=FT\{h(t)\}=\int A\mathrm{e}^{-t/\tau}u(t)\mathrm{d}t=\int_0^{\infty}A\mathrm{e}^{-t/\tau}\mathrm{e}^{-\mathrm{j}\omega t}\mathrm{d}t=\frac{A}{\mathrm{j}\omega+1/\tau}$$

因此，频域的细胞膜输出电压是

$$V(\omega)=H(\omega)I/(\omega)=\frac{A}{\mathrm{j}\omega+1/\tau}\frac{1}{\mathrm{j}\omega}=\left(\frac{A}{\mathrm{j}\omega}-\frac{A}{\mathrm{j}\omega+1/\tau}\right)\tau$$ ■

10.6.4 模拟滤波器

滤波器是一类特殊的线性系统，广泛用于生物信号的处理。理论上，滤波器能够选择性地去除不想要的信号成分，保留或加强某些感兴趣的信号成分。尽管很少有人留意，但是，日常生活中其实到处存在各种各样的滤波器。例如，太阳镜就是一种滤波器，它可以滤除太阳光中的紫外线，以便尽可能减小晒伤和患皮肤癌的危险。音响设备中也有很多滤波器，其高音控制和低音控制就是特殊的滤波器，用户可以用它们来增强或者抑制高频声音（高音）和低频声音（低音）的含量，从而获得所需的音质。

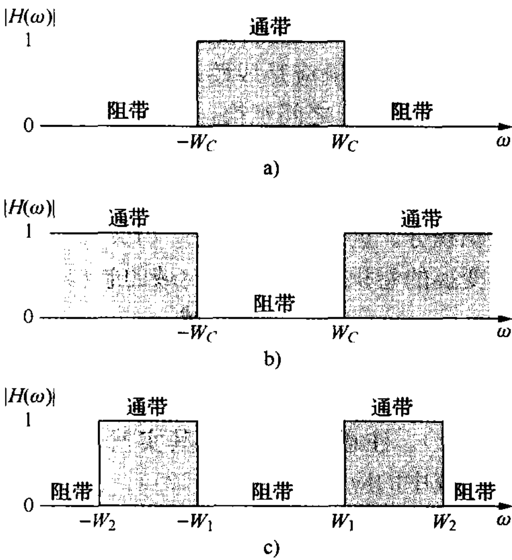


图 10.16 各种理想滤波器的频域幅值响应特性 $|H(\omega)|$

a）低通滤波器 b）高通滤波器 c）带通滤波器

注：输出信号中保留了阴影区域的通带信号，同时有选择地去除阻带内的信号。

滤波器在生物信号分析中具有重要的作用，其原因之一是生物医学仪器所测得的信号往往包含了不想要的噪声，这些噪声会使我们所感兴趣的信号波形产生失真，从而很难作出可靠的诊断。如果可以去除这些噪声，那么，就可以显著提高信号的质量，从而使误诊的可能性降为最小。

根据频谱的不同，滤波器可以分为低通、高通和带通三大类。低通滤波器可以将信号中的高频分量去除，有选择地让低频分量通过，因此称为“低通”（见图 10.16a）。有时，这种滤波器在去除高频分量的同时还可以进一步放大低频分量。高通滤波器则与低通滤波器恰好相反（见图 10.16b），它可以有选择地通过高频分量而去除低频分量。音响系统中的高音控制就是一种增强高频分

量的高通滤波器，因此会产生尖脆而饱满的声音。相反，低音控制则是一种选择性地增强低频分量的低通滤波器，从而产生一种“温和”的音质。带通滤波器可以看作低通滤波器和高通滤波器的结合，它同时把低频和高频分量都去掉，只留下一部分“窄带”频率分量（见图 10.16c），因此称为“带通”。简单地把低通滤波器和高通滤波器级联起来，就可以形成带通滤波器（参见本书第 8 章）。

滤波器是线性系统，在时域上，其输出等于输入与滤波器冲激响应的卷积（见式（10.32））；在频域上，其输出等于滤波器传递函数与输入的傅里叶变换的乘积（见式（10.35））。图 10.17 所示是理想低通滤波器的冲激响应及其传递函数，该传递函数在通带内的值为 1，在阻带内的值为 0。由于线性系统的输出在频域上等于输入的傅里叶变换与传递函数的乘积，即 $Y(\omega)=X(\omega)H(\omega)$，并且此滤波器通带内的频率分量都乘以 1，因此，通带内的信号都可以无改变地通过，直接成为输出。相反，阻带内的频率分量乘以 0，因此，在滤波器的输出中这些分量都被去除。

在时域中，理想模拟低通滤波器的冲激响应是

$$h_{\text{LP}}(t)=\frac{W_c}{\pi}\text{sinc}(W_c t) \tag{10.36}$$

式中　W_c——滤波器的截止频率，$W_c=2\pi f_c$。

在频域中，理想滤波器的传递函数是

$$H_{\text{LP}}(\omega)=\begin{cases}1 & |\omega|<W_c\\ 0 & |\omega|>W_c\end{cases} \tag{10.37}$$

图 10.17 显示了这种理想低通滤波器的时域和频域的双重表示法。注意，传递函数只在通带内取值为 1，在截止频率处，滤波器的传递函数从通带的 1 变为阻带 0。

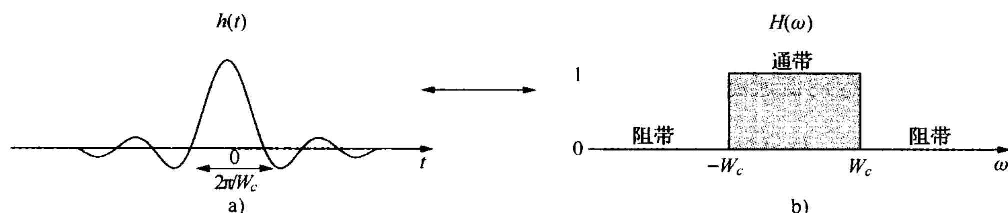


图 10.17　理想低通滤波器的时域和频域表示

a）时域的冲激响应 $h(t)$　b）频域的传递函数 $H(\omega)$

在频域中，理想高通滤波器的性能恰好与低通滤波器相反，即

$$H_{\text{HP}}(\omega)=\begin{cases}0 & |\omega|<W_c\\ 1 & |\omega|>W_c\end{cases} \tag{10.38}$$

也就是说，通带在截止频率以上，而阻带在截止频率以下。因此，这种滤波器保留信号的高频分量，有选择地去除低频分量。理想高通滤波器的传递函数很容易从理想低通滤波器导出，即

$$H_{\text{HP}}(\omega)=1-H_{\text{LP}}(\omega) \tag{10.39}$$

在时域，理想高通滤波器的冲激响应就是

$$h_{HP}(t)=\delta(t)-h_{LP}(t)=\delta(t)-\frac{W_c}{\pi}\text{sinc}(W_c t) \tag{10.40}$$

图 10.16b 所示就是理想高通滤波器的传递函数形状。

这里要介绍的最后一类滤波器是带通滤波器。带通滤波器比低通和高通滤波器都要复杂，它有低频截止频率 W_1 和高频截止频率 W_2。图 10.16c 显示了理想带通滤波器的传递函数幅值。只有处于两个截止频率之间的信号可以通过滤波器，所有其他信号都被去除。理想带通滤波器的传递函数是

$$H_{BP}(\omega)=\begin{cases}1 & W_1<|\omega|<W_2\\ 0 & \text{其他}\end{cases} \tag{10.41}$$

在频域上，将截止频率为 W_1 的高通滤波器和截止频率为 W_2 的低通滤波器联合起来，就可以得到理想带通滤波器。因此，带通滤波器的传递函数可以表示为低通滤波器和高通滤波器传递函数的乘积，即

$$H_{BP}(\omega)=H_{HP}(\omega)\cdot H_{LP}(\omega) \tag{10.42}$$

该滤波器传递函数的傅里叶反变换就是带通滤波器时域的冲激响应，即

$$h_{BP}(t)=h_{HP}(t)*h_{LP}(t) \tag{10.43}$$

对于式（10.42）应用卷积定理（式（10.14）和式（10.15）），就可以得到上式。

例 10.21 肌电信号（EMG）所包含的频率范围为 25～100 Hz，请设计一个滤波器去除 EMG 中其他不需要的噪声。

解： 本例题需要设计一个通带为 25～100 Hz 的带通滤波器。首先计算滤波器的两个截止频率，单位为 rad/s。由于 $W_c=2\pi f_c$，因此

$$W_1=50\pi$$
$$W_2=200\pi$$

然后，确定相应的低通滤波器和高通滤波器的冲激响应，即

$$h_{HP}(t)=\delta(t)-\frac{W_1}{\pi}\text{sinc}(W_1 t)=\delta(t)-50\text{sinc}(50\pi t)$$

$$h_{LP}(t)=\frac{W_2}{\pi}\text{sinc}(W_2 t)=200\text{sinc}(200\pi t)$$

于是，带通滤波器的冲激响应就是

$$h_{BP}(t)=h_{BP}(t)*h_{LP}(t)=[\delta(t)-50\text{sinc}(50\pi t)]*200\text{sinc}(200\pi t)$$ ■

上述理想模拟滤波器是各种实际滤波器设计的理论基础。由于理想滤波器的冲激响应是无限长的（$-\infty\sim\infty$），其输出的形成需要无限长时间，因此，实际的模拟滤波器不可能严格达到理想滤波器的性能指标。通常，模拟滤波器用电子电路实现，它们可以设计成各种接近理想低通、高通和带通的滤波器来满足不同的应用需求，包括处理各种生物医学信号。

10.6.5 数字滤波器

就像模拟系统要用微分方程来描述一样，数字系统要用差分方程来描述。差分

方程实质上是微分方程以一定的采样频率采样得到的离散化形式。实时数字滤波器（即差分方程）的一般形式是

$$y(k)=\sum_{m=0}^{M}b_m x(k-m)-\sum_{m=1}^{N}a_m y(k-m) \tag{10.44}$$

式中　$x(k)$ 和 $y(k)$ ——离散化的输入序列和输出序列。

例如，如果 $M=2$ 且 $N=2$，那么

$$y(k)=b_0x(k)+b_1x(k-1)+b_2x(k-2)-a_1y(k-1)-a_2y(k-2)$$

其中 $x(k)$ 和 $y(k)$ 分别为 k 时刻的输入和输出，$x(k-1)$ 和 $y(k-1)$ 分别为前一个采样点时刻的输入和输出，同理，$x(k-2)$ 和 $y(k-2)$ 分别为前两个采样点时刻的输入和输出。

数字系统与模拟系统一样，也可以用冲激响应 $h(k)$ 和卷积来定义（见式(10.33)）。如果冲激响应具有有限个非零数据点，则该滤波器被称为有限冲激响应（Finite Impulse Response，FIR）滤波器。如果冲激响应有无限多个非零数据点，则该滤波器被称为无限冲激响应（Infinite Impulse Response，IIR）滤波器。数字滤波器的一个优点是可以方便地计算任意输入所对应的输出响应。

例 10.22　请计算如下数字滤波器的冲激响应

$$y(k)=\frac{1}{2}x(k)+\frac{1}{2}y(k-1)$$

解： 假设输入开始之前系统的输出均为 0，即对于 $n<0$，$y(n)=0$，则

$$y(-2)=\frac{1}{2}\delta(-2)+\frac{1}{2}y(-3)=0+0=0$$

$$y(-1)=\frac{1}{2}\delta(-1)+\frac{1}{2}y(-2)=0+0=0$$

$$y(0)=\frac{1}{2}\delta(0)+\frac{1}{2}y(-1)=\frac{1}{2}+0=\frac{1}{2}$$

$$y(1)=\frac{1}{2}\delta(1)+\frac{1}{2}y(0)=0+\frac{1}{2}\left(\frac{1}{2}\right)=\left(\frac{1}{2}\right)^2$$

$$y(2)=\frac{1}{2}\delta(2)+\frac{1}{2}y(1)=0+\frac{1}{2}\left(\frac{1}{2}\right)^2=\left(\frac{1}{2}\right)^3$$

$$y(3)=\frac{1}{2}\delta(3)+\frac{1}{2}y(2)=0+\frac{1}{2}\left(\frac{1}{2}\right)^3=\left(\frac{1}{2}\right)^4$$

$$\cdots$$

$$y(k)=\left(\frac{1}{2}\right)^{k+1}u(k)$$

该滤波器的冲激响应是一个指数序列，由于其冲激响应无限长，这是一个 IIR 滤波器。■

例 10.23　请计算如下数字滤波器的冲激响应

$$y(k)=\frac{1}{3}x(k)+\frac{1}{3}x(k-1)+\frac{1}{3}x(k-2)$$

解： 假设输入开始之前系统的输出均为 0，即对于 $n<0$，$y(n)=0$，则

$$y(-2)=\frac{1}{3}\delta(-2)+\frac{1}{3}\delta(-3)+\frac{1}{3}\delta(-4)=0+0+0=0$$

$$y(-1)=\frac{1}{3}\delta(-1)+\frac{1}{3}\delta(-2)+\frac{1}{3}\delta(-3)=0+0+0=0$$

$$y(0)=\frac{1}{3}\delta(0)+\frac{1}{3}\delta(-1)+\frac{1}{3}\delta(-2)=\frac{1}{3}+0+0=\frac{1}{3}$$

$$y(1)=\frac{1}{3}\delta(1)+\frac{1}{3}\delta(0)+\frac{1}{3}\delta(-1)=0+\frac{1}{3}+0=\frac{1}{3}$$

$$y(2)=\frac{1}{3}\delta(2)+\frac{1}{3}\delta(1)+\frac{1}{3}\delta(0)=0+0+\frac{1}{3}=\frac{1}{3}$$

$$y(3)=\frac{1}{3}\delta(3)+\frac{1}{3}\delta(2)+\frac{1}{3}\delta(1)=0+0+0=0$$

$$y(4)=\frac{1}{3}\delta(4)+\frac{1}{3}\delta(3)+\frac{1}{3}\delta(2)=0+0+0=0$$

$$\cdots$$

$$y(k)=0;k\geqslant 3$$

这是一个只有 3 个非零数值的 FIR 滤波器。 ■

IIR 滤波器在仿真模拟系统时特别有用。IIR 滤波器的主要优点是实现滤波所需的滤波器系数比 FIR 滤波器少，也就是 IIR 滤波器通常效率更高。IIR 滤波器的主要缺点是信号可能会变形。而 FIR 滤波器可以设计成对称的，从而防止信号的变形。变形问题的处理方法本书不做介绍。

顾名思义，数字滤波器常用于实现特定的滤波功能，如低通滤波器、高通滤波器、带通滤波器、带陷滤波器等。不过，数字滤波器也可以用于仿真模拟系统，如微分系统和积分系统等。有关数字滤波器设计的教科书有很多，本书下面只介绍其中的主要内容。

从数字滤波器到传递函数

通过重组差分方程（10.23），并应用式（10.21），可以得到数字系统的传递函数 $H(z)$。$H(z)$ 是输出信号的 Z 变换 $Y(z)$ 除以输入信号的 Z 变换 $X(z)$ 所得到的商，即

$$y(k)+a_1y(k-1)+a_2y(k-2)\cdots+a_Ny(k-N)=b_0x(k)+b_1x(k-1)+\cdots+b_Mx(k-M)$$

$$Y(z)+a_1z^{-1}Y(z)+a_2z^{-2}Y(z)\cdots+a_Nz^{-N}Y(z)=b_0X(z)+b_1z^{-1}X(z)+b_2z^{-2}X(z)\cdots+b_Mz^{-M}X(z)$$

$$Y(z)(1+a_1z^{-1}+a_2z^{-2}\cdots+a_Nz^{-N})=X(z)(b_0+b_1z^{-1}-b_2z^{-2}\cdots+b_Mz^{-M})$$

$$H(z)=\frac{Y(z)}{X(z)}=\frac{b_0+b_1z^{-1}+b_2z^{-2}\cdots+b_Mz^{-M}}{1+a_1z^{-1}+a_2z^{-2}\cdots+a_Nz^{-N}} \tag{10.45}$$

从传递函数到频率响应

数字系统的频率响应 $H'(\Omega)$ 可以直接由 $H(z)$ 计算得到，其中 Ω 的单位是弧

度。如果如前所述数据是由模拟信号采样得到，那么，ω 与 Ω 之间的关系是 $\Omega=\omega T$，则

$$H'(\Omega)=H(z)\big|_{z=e^{j\Omega}} \tag{10.46}$$

对于线性系统，以下输入序列

$$x(k)=A\sin(\Omega_0 k+\Phi)$$

所产生的输出稳态序列将为

$$y(k)=B\sin(\Omega_0 k+\Phi)$$

其中，B 和 Φ 的值可以直接计算，即

$$B=A|H'(\Omega_0)|,\varnothing=\Phi+(H'(\Omega_0))\text{相位角}$$

例 10.24　假设例 10.22 数字滤波器的输入序列为

$$x(k)=100\sin\left(\frac{\pi}{2}k\right)$$

那么，其输出的稳态序列是什么？

解：首先将差分方程

$$y(k)-\frac{1}{2}y(k-1)=\frac{1}{2}x(k)$$

变换到 z 域，即：

$$Y(z)-\frac{1}{2}Y(z)z^{-1}=Y(z)\left[1-\frac{1}{2}z^{-1}\right]=\frac{1}{2}X(z)$$

求解 $H(z)$，得到滤波器的传递函数，即

$$H(z)=\frac{Y(z)}{X(z)}=\frac{\frac{1}{2}}{1-\frac{1}{2}z^{-1}}$$

然后，为了求取输出信号，就要计算输入正弦波频率($z=e^{j\frac{\pi}{2}}$)上的传递函数，即

$$H'\left(\frac{\pi}{2}\right)=H(e^{j\frac{\pi}{2}})=\frac{\frac{1}{2}}{1-\frac{1}{2}e^{-j\frac{\pi}{2}}}=\frac{\frac{1}{2}}{1+\frac{1}{2}j}=0.4-j0.2=0.45e^{-j0.15\pi}$$

由该传递函数可知，将输入信号的幅值乘以 0.45 并加上一个相位因子 -0.15π 弧度，就可以得到所需的输出。因此，输出为

$$y(k)=45\sin\left(\frac{\pi}{2}k-0.15\pi\right)$$ ■

设计滤波器时，首先要确定信号中需要保留的频率分量和需要去除的频率分量。对于理想滤波器 $|H'(\Omega_{\text{保留}})|=1$ 且 $|H'(\Omega_{\text{去除}})|=0$。例 10.23 和例 10.24 中的两个实际滤波器都属于低通滤波器，如图 10.18 所示，它们的频率响应都不太好。图中添加了截止频率为 $\pi/4$ 的理想低通滤波器作为比较，其频率响应为 $|\Omega|<\pi/4$ 时 $|H'(\Omega)|=1$，而 $|\Omega|>\pi/4$ 时 $|H'(\Omega)|=0$。

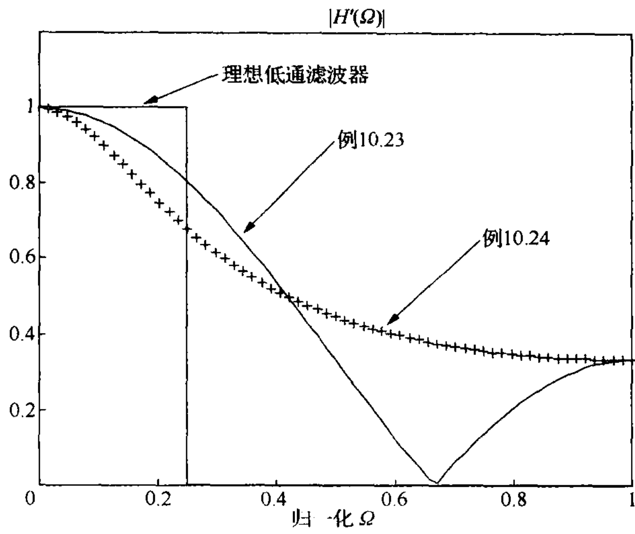


图 10.18 例 10.23 和例 10.24 中所述低通滤波器的频域特性

注：图中添加了截止频率为 $\pi/4$ 弧度（以 π 弧度归一化之后为 0.25）的理想低通滤波器作为对比。低通滤波器的截止频率通常定义为归一化幅值等于 $1/\sqrt{2}$（即约为 0.71）处的频率。两个数字滤波器在 f_{max}（即归一化的 $\Omega=1$）处的幅值相等。

10.7 信号平均

生物信号的测量经常会受到噪声的干扰，使信号的特征难以识别，影响临床诊断的可靠性。

测量信号可以用理想的无噪信号 $x(t)$ 与独立的噪声 $n(t)$ 之和来表示，即

$$x_i(t)=x(t)+n(t) \tag{10.47}$$

式中 $x_i(t)$ ——第 i 次实验所记录的信号。

注意，第 i 次记录中既包含确定性信号 $x(t)$ 也包含随机噪声 $n(t)$。其中，确定性信号在每次实验记录中固定不变，而噪声却具有内在的变化，它可能由多个不同的噪声源产生。由于各次实验记录中随机噪声 $n(t)$ 都不同，因此，每次重复记录的信号都变化明显。例如，心电图（ECG）测量电极会同时记录到来自肌肉、肺部呼吸的信号，还会记录到仪器内部电子噪声等干扰，如来自电源的 60 Hz 工频噪声等。这些信号与心脏搏动无关，都是测量信号中的噪声。其他某些难以预料的心脏搏动变化，比如喝过浓咖啡之后兴奋引起的心搏变化，也被认为是噪声。

我们已经知道，设计合适的滤波器可以滤除噪声，分离出所需信号，使信号变得比较干净，从而提高信号质量和临床诊断的可靠性。如果在频域中噪声分量和信号分量不重叠，那么，只要设计一个滤波器，保留或增强想要的信号 $x(t)$，同时去除不想要的噪声 $n(t)$ 就可以了。这种去噪方法简单有效，但是在很多情况下生物信

号和噪声的频谱是重叠的，这种方法不能用。

很多生物信号本质上接近周期信号，例如，与心脏搏动相关的血压、血流速度、心电图等都属于这类信号。但是，由于呼吸等噪声的影响，各次心跳之间总会有所差异。如图 10.2 所示的血压信号就包含了上述各种噪声。

血压信号含有很多特征，医生和研究人员利用这些特征可以确定病人的健康状况。常用的血压测量参数有：心脏射血期（即收缩期）的最高血压、主动脉瓣关闭时（即舒张期）达到的最低血压、收缩期早期的最大血压变化率（它表示心脏收缩的强度）以及舒张期血压按指数衰减的时间常数（它可以反映血管阻抗和顺应性）。求取这些参数的一种方法就是先计算一连串心脏搏动中每次心跳的参数值，然后求它们的平均值。这种方法通常不可行，因为单次测量中所含的噪声会使某些参数的精确计算难以实现。还有一种方法是先计算所有记录的叠加平均信号，得到典型的心脏搏动信号之后，再根据这个平均波形计算各个参数值。也就是

$$\bar{x}(t) = \frac{1}{N}\sum_{i=1}^{N} x_i(t) \tag{10.48}$$

如果信号是离散的，那么上式变为

$$\bar{x}(k) = \frac{1}{N}\sum_{i=1}^{N} x_i(k) \tag{10.49}$$

其中，连续信号 $x_i(t)$ 和离散信号 $x_i(k)$ 表示总共 N 次记录中的第 i 次心脏搏动数据。信号 $\bar{x}(t)$ 和 $\bar{x}(k)$ 则是均值，也就是叠加平均计算之后的平均波形。将式 (10.47) 代入式 (10.48) 可得

$$\bar{x}(t) = x(t) + \frac{1}{N}\sum_{i=1}^{N} n(t) = x(t) + \varepsilon(t) \tag{10.50}$$

如果噪声 $n(t)$ 是纯随机信号，那么上式所包含的测量误差 $\varepsilon(t)$ 在 $N\to\infty$时趋于 0，因此，对于非常大的 N 值，$\varepsilon(t)$ 会趋于 0，从而使 $\bar{x}(t)\approx x(t)$。这是一个很有用的方法！它意味着通过多次测量信号的简单平均可以有效地去除噪声。实质上，如果记录次数足够多，平均信号就能够接近真正的无噪信号波形。

由于心脏搏动的平均波形很接近真实信号，因此，根据这种平均波形计算各个参数指标就比较精确。很多生物信号采集系统都具备求平均信号的功能，各次叠加求和过程由信号本身或者与信号相关的某个特征量触发启动，含有特征波峰的 ECG 信号就常用作与心搏相关的记录数据的叠加平均触发信号。图 10.19 所示是图 10.2 血压数据的叠加平均波形。图 10.20 显示了听觉脑干反应（Auditory Brainstem Response，ABR）EEG 记录信号的平均过程。

上述两个例子说明了时域信号的平均叠加过程。对于随机信号有时用频域的平均叠加会更好。如图 10.21 所示的脑电图（EEG）就属于随机信号。图中显示了病人头部枕叶上采集的一段 EEG 信号，采样频率为 16 kHz。人在睡眠、静息和清醒等不同状态下，EEG 信号具有不同的频率含量，因此，EEG 的分析通常在频域进行。计算不同频段上 EEG 信号的功率就是一种常用的分析方法，信号功率可以用傅里叶变换的二次方来估计。

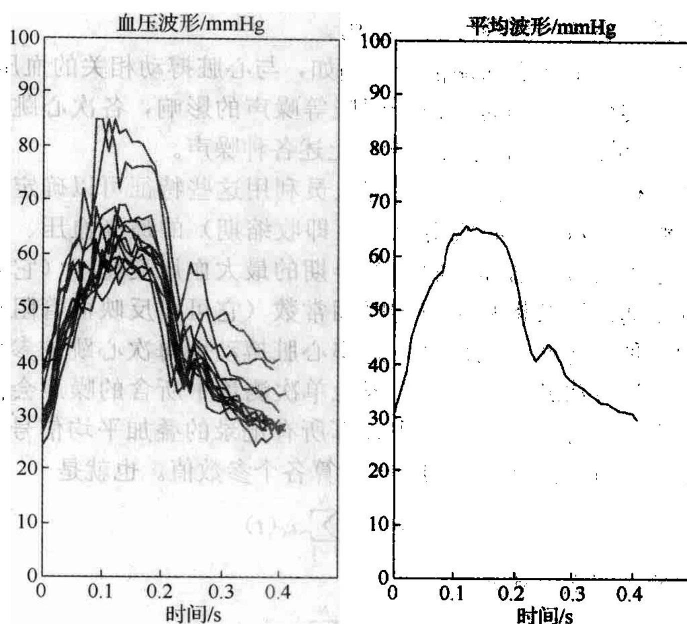


图 10.19 图 10.2 所示的血压波形记录数据的叠加平均

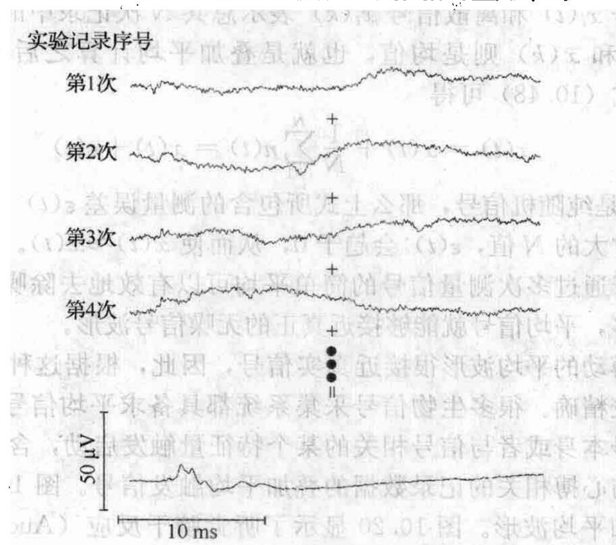


图 10.20 在颞叶上测得的 0 时刻发生的短声音脉冲引起的听觉诱发响应重复实验记录信号

注：各个单次记录信号中的听觉响应被随机噪声淹没（图中只显示了 1000 次重复实验中的前 4 次记录）。从图中下方显示的 1000 次记录的叠加平均波形上清晰可见听觉响应信号。

如果用 DFT 来计算信号各个频率分量上的功率，那么，由于每段记录信号的功率都有变化，计算所得的各段信号的功率估计值也有很大变化。噪声与测量信号的

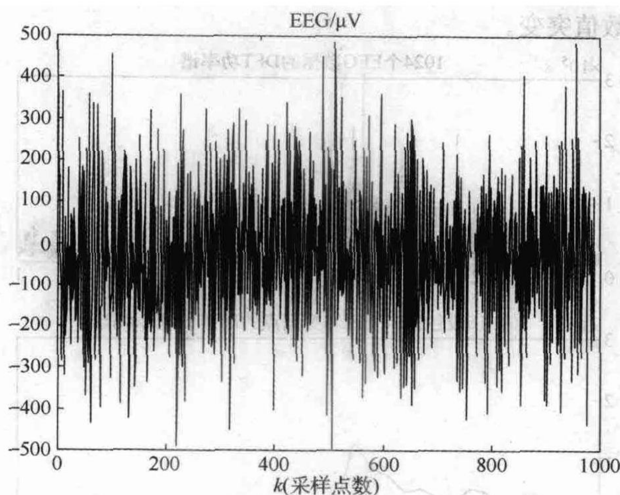


图 10.21　枕叶上记录的包含 1000 个采样点的 EEG 信号
（采样频率为 16 kHz）

大小相当。为了减小噪声的方差，就要采用统计学的方法。一种常用的信号功率估计方法是 Welch 法，也称为周期图平均法。这种方法先将信号分为 L 段（尽可能不要交叠），每段包含 N 个点；然后对 L 段中的每一段进行 DFT 功率计算；最后，将这 L 段的功率数据求平均，所得结果就是 N 个频率点上的平均功率估计值。算法如下。

设第 i 段的 N 个数据点为

$$x_i(k)=x(k+(i-1)N)\quad 0\leqslant k\leqslant N-1,1\leqslant i\leqslant L$$

如果各段信号之间连续且不交叠，那么，单独第 i 段数据 DFT 计算得到的功率估计值为

$$\hat{P}_i(m)=\frac{1}{N}\left|\sum_{k=0}^{N-1}x_i(k)\mathrm{e}^{-\mathrm{j}\frac{2\pi mk}{N}}\right|^2\quad 其中\ 0\leqslant m\leqslant N-1\tag{10.51}$$

这里用 m 表示功率值所处的频率是 $\Omega=2\pi m/N$ 弧度。信号的平均功率谱就是每个频率点上的平均，即

$$\hat{P}(m)=\frac{1}{L}\sum_{i=1}^{L}\hat{P}_i(m)\tag{10.52}$$

由于 N 决定了频域的分辨率，因此 N 的选择非常重要。例如，如果数据的采样频率为 500 Hz，并且所需频率分辨率为 1 Hz，那么，L 段数据的每段长度至少要 1 s，也就是包含 500 个采样点（即 $N=500$）。如果频率分辨率只要 10 Hz 就足够了，那么，每段长度只要 0.1 s，也就是 50 个数据采样点。这种平均处理可以将信号功率谱估计的方差降为 $1/L$。图 10.22 说明了 EEG 信号的这种平均处理方法。该方法还有很多改进形式，例如，如果所需 L 值比较大，但是可用的数据采样点却不够多，那么，数据段之间可以交叠；另外，还可以将每个数据段乘以一个窗函数，以避免数据段

起始和末尾处的数值突变。

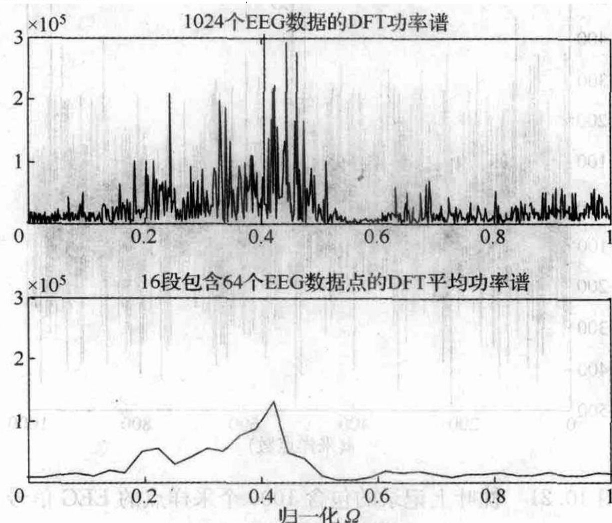


图 10.22 EEG 信号的 DFT 平均功率谱

注：上图是用整个原始记录信号计算得到的 DFT 功率谱；下图是将记录信号分为包含 64 个数据点的 16 段数据之后计算得到的周期图平均 DFT 功率谱。

例 10.25 假设某个受到随机噪声 $n(k)$ 干扰的正弦信号是

$$x(k)=\sin(\pi/4k)+n(k)$$

请编写 MATLAB 程序，说明叠加平均法可以去除噪声并显示出确定性信号。请分别作图表示 1 段、10 段、100 段数据的平均结果。

解：MATLAB 程序如下：

```
k=1:64; %Discrete Time Axis
for i=1:100 %Generating 100 signal Trials
x(i,:)=sin(pi/4*k)+randn(1,64); %i-th trial
end
X1=x(1,:); %1 Averages
X10=mean(x(1:10,:)); %10 Averages
X100=mean(x); %100 Averages
subplot(311) %Plotting Results, 1 Average
plot(k,X1,'k')
axis([1 64 -3 3])
title('1 Average')
ylabel('Amplitude')
subplot(312) %Plotting Results, 10 Averages
plot(k,X10,'k')

axis([1 64 -3 3])
title('10 Averages')
```

```
ylabel('Amplitude')
subplot(313) %Plotting Results, 100 Averages
plot(k,X100, 'k')
axis([1 64 -3 3])
title('100 Averages')
xlabel('Discrete Time')
ylabel('Amplitude')
```

该程序的运行结果如图 10.23 所示。 ■

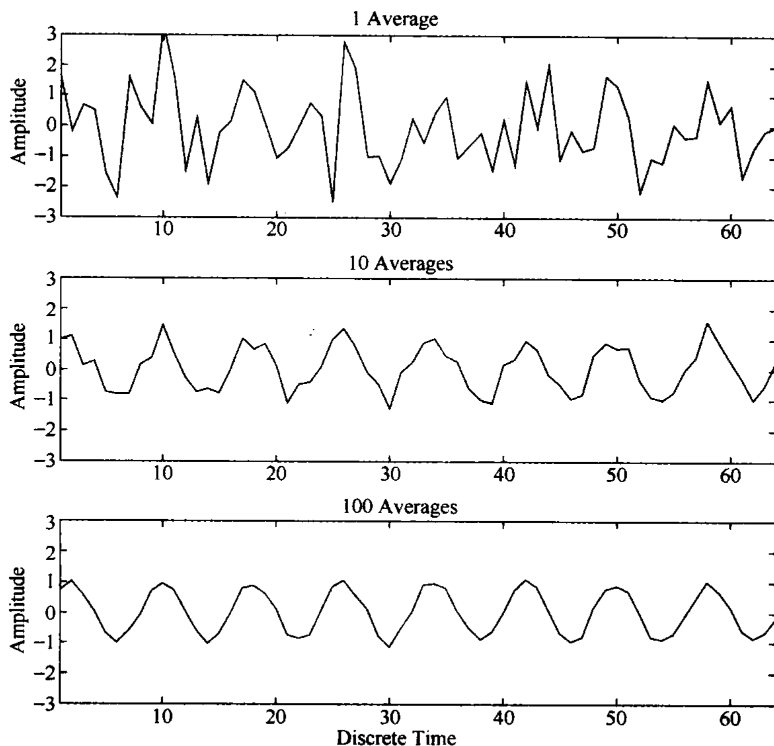


图 10.23　例 10.25 的 MATLAB 程序运行结果

注：用叠加平均法去除含噪正弦波信号中的噪声，图中所示分别是 1 段、10 段和 100 段数据的平均。

10.8　小波变换和短时傅里叶变换

式（10.6）表示的傅里叶变换是一种常用的信号处理方法，它将信号分解为不同频率的正弦波分量。很多应用中，特别是对于随时间变化不大的平稳信号，由傅里叶变换求得的信号总体的频率分布基本上包含了所需的信息。但是，傅里叶变换不能描述信号随时间变化的过程。

短时傅里叶变换（Short-Time Fourier Transform，STFT）和小波变换（Wavelet

Transform，WT）在求取信号频域特性的同时还可以保留信号的时域信息。STFT 每次只对一小段数据（即数据窗）进行傅里叶变换，这样就可以在时间和频率构成的二维空间中描述信号的时频特性。该变换的数学表示是

$$X(\omega,a)=\int_{-\infty}^{\infty}x(t)g(t-a)e^{-j\omega t}dt \tag{10.53}$$

式中 $g(t)$ ——窗函数，它可以是简单的方波函数。

STFT 的反变换是

$$X(\omega,a)=K_g\iint X(\omega,a)g(t-a)e^{j\omega t}dtda \tag{10.54}$$

式中 K_g——所用的数据截取窗函数。

为了避免数据截取时数据段两端突变引起的频谱的旁瓣振荡效应，矩形截取窗可以改为两端渐变的截取窗。图 10.24 显示了这两种截取窗，两个窗叠加在严格重复的主动脉血压周期信号上。为了能够看得清楚，窗的幅值放大了 100 倍。

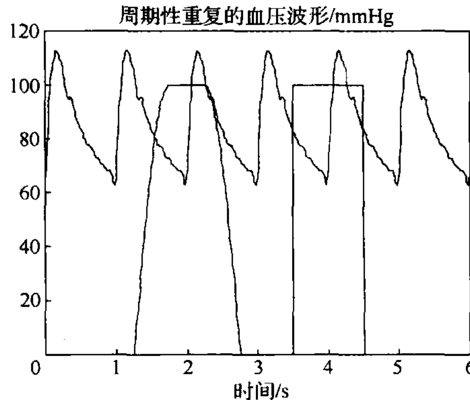


图 10.24 严格周期性重复的主动脉血压波形以及可以用于 STFT 计算的两种窗函数

注：窗宽都约为一个血压周期的宽度。两端渐变的窗（左）有利于避免两端突变的窗（右）所引起的频谱旁瓣振荡效应。为了能够看得清楚，窗的幅值放大了 100 倍。

图 10.25 显示了用 3 种不同窗宽的矩形窗截取图 10.24 所示周期信号计算得到的 STFT 幅值频谱，这 3 种窗宽分别为 1/2 个周期、1 个周期和 2 个周期。其中，上图中的垂直线条反映了信号中含有比窗宽更大的周期信息。下面两个图中的彩色水平实线表示在所用窗宽下这些频率成分与时间完全无关。由于此时窗宽包含了 1 个或者 2 个完整的周期，应该出现这样的结果。下图中水平暗线与亮线交替出现，表明数据窗内包含多个重复周期，暗线表示相应的频率分量很小或者没有。

作为对比，图 10.26 显示了图 10.2 的非周期性血压波形的 STFT 幅值频谱，所用的窗宽尽可能接近心率周期。信号中的均值已去除，因此最低的 0 频率水平线处的数值变化反映的是呼吸引起的变化。心率频率处（即“1”频率处）的频谱随时间变化最小，以后，随着频率的增大，变化也增大。

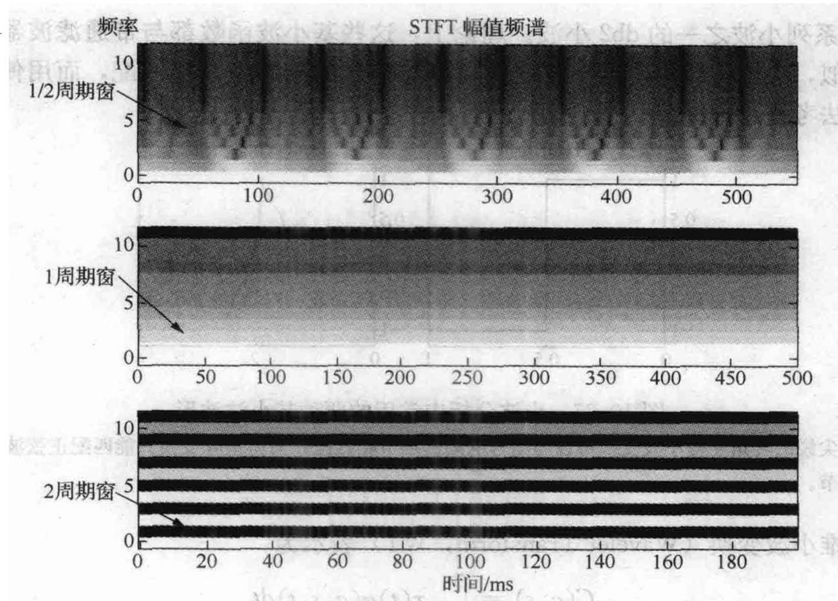


图 10.25　图 10.24 中严格重复的周期信号的 STFT 二维幅值频谱
(其中 3 个图所用的窗宽分别为 1/2 个周期、1 个周期和 2 个周期)

注：颜色越亮表示频谱幅值越大。例如，位于 0 频率的条带对应于均值，在 3 个图中其幅值都是最大。0 以上的条带对应于数据中的谐波分量，通常随着频率的增大而减小。

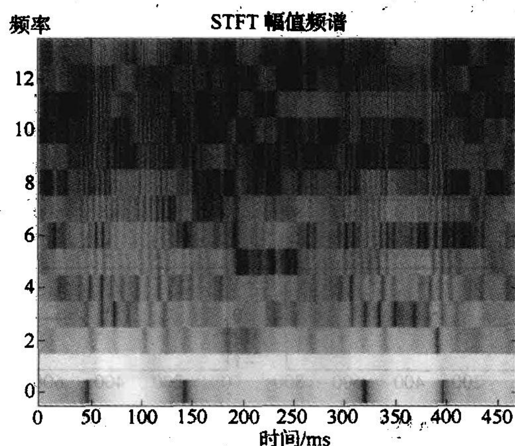


图 10.26　图 10.2 的非周期性血压波形的 STFT 二维幅值频谱

STFT 主要的缺点是在整个计算过程中窗的宽度固定不变。小波分析法与 FT 和 STFT 的区别就在于：其信号的分解不再限于正弦波，并且窗宽是可变的。在小波分析中，信号被分解为带位移因子和尺度因子的各个基小波 $\Psi(t)$ 的组合，基小波也称为“母”小波。图 10.27 所示是两个基小波的例子，左边是 Haar 小波，右边是 Dau-

bechies 系列小波之一的 db2 小波。理论上，这些基小波函数都与带通滤波器的冲激响应类似，尖锐的转角使得小波变换可以与信号的局部细节相匹配，而用傅里叶变换就无法考察这种细节。

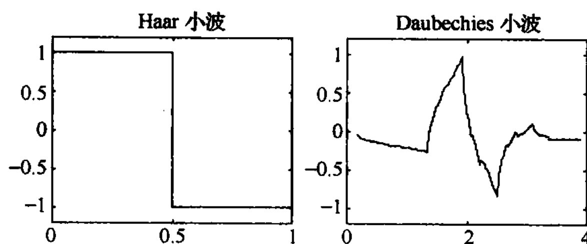


图 10.27 小波分析中常用的两种基小波波形

注：尖锐的转角使得小波变换可以与信号的局部细节相匹配，而傅里叶变换只能匹配正弦波，无法考察这种细节。

二维小波变换（Wavelet Transform，WT）表示为

$$C(a,s)=\int_{-\infty}^{\infty}x(t)\varphi(a,s,t)\mathrm{d}t \tag{10.55}$$

式中 a——尺度因子；

s——位移因子；

C——可以理解为带尺度因子和位移因子的各种小波与信号数据之间的相关系数。

图 10.28 显示了具有不同尺度和不同位移的 db2 小波函数 $\varphi(t)$，例如，$\varphi(2,-100,t)=\varphi(2t-100)$。

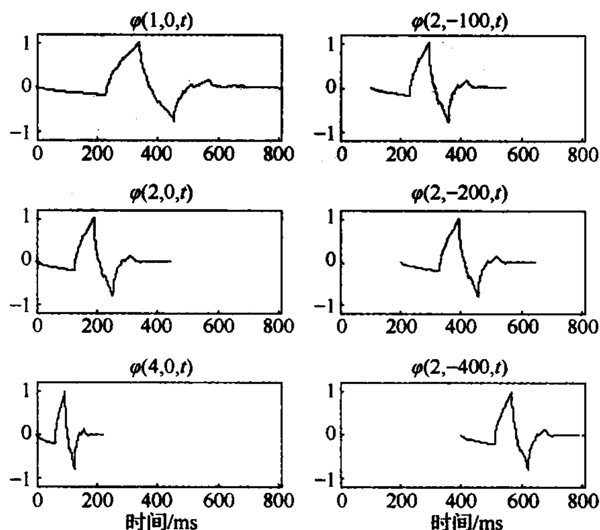


图 10.28 几种不同尺度和位移下的 db2 小波

注：左上方是基小波 $\varphi(t)$。各图上方标记的含义是 φ（尺度，位移，t）。因此 $\varphi(1,0,t)=\varphi(t)$，$\varphi(2,-100,t)=\varphi(2t-100)$，以此类推。

二维小波反变换为

$$x(t) = K_{\varphi}\iint C(a,s)\varphi(a,s,t)\mathrm{d}t\mathrm{d}s \tag{10.56}$$

可用于从小波系数 $C(a, s)$ 中重建原信号 $x(t)$，其中 K_{φ} 是所用小波函数。

实际应用中，小波分析只是计算某一组离散尺度和位移下的变换，这称为离散小波变换（参见 MATLAB 的小波工具箱）。其中一种算法是反复用数字低通滤波器和高通滤波器将信号分解为低频分量（即“大尺度”下的“近似”分量）和高频分量（即“小尺度”下的“细节”分量），这些滤波器都是基小波的函数。每个滤波器的输出数据点数与输入相同。为了使最后得到的不同尺度下的数据总数保持不变，每个滤波器输出序列都每隔一个数据点保留一个数据，另一个数据则被删除，这个过程称为降采样。利用这种降采样数据和另一组“重建”滤波器，就可以求得小波反变换，重建原信号。显然，离散小波变换的反变换确实存在！

利用上述算法可以快速求得小波变换系数。离散小波分析的重要作用是可以在不同尺度分辨率下考察信号波形，从而能够选择信号中有用的“近似”分量和“细节”分量来重建所需的波形。其应用包括：检测信号中的不连续点和断点，检测长时程的信号变化，检测自相似性（如分形树等），识别单纯频率成分（这与傅里叶变换相似），以及滤波、去噪和信号压缩等。

作为对比，下面用离散傅里叶变换和离散小波变换分别处理图 10.2 所示的血压波形。图 10.29 显示了整个数据段的 DFT 频谱，各个心率谐波处具有一定宽度的波峰以及不规则的频谱幅值，反映了心脏搏动之间的差异，较低的频率处还可看到呼吸节律。

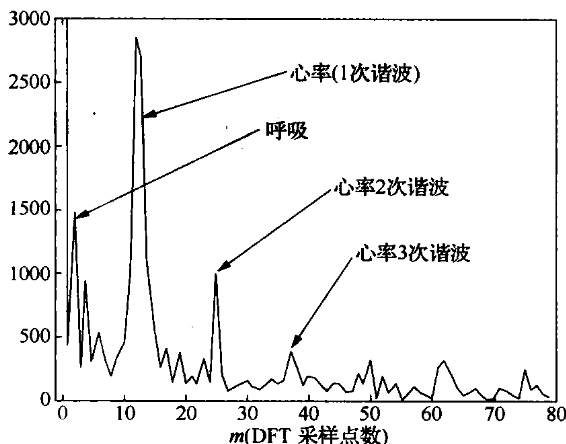


图 10.29　图 10.2 所示血压波形的 DFT 频谱（心率的 1 次、2 次和 3 次谐波清晰可见）

再来看利用 MATLAB 小波工具箱的小波变换函数处理上述血压数据的结果。图 10.30 是小波变换系数的二维显示，x 轴是位移，y 轴是尺度，小尺度在下，大尺度在上。最大尺度处明显地显示出信号中所包含的两个呼吸周期。除了小波变换系数之外，更重要的是可以选择信号的不同细节分量和近似分量。当尺度从 a1 变化到 a7 时，近似分量从心率成分为主变化到呼吸成分为主。从接近心率的小尺度细节分量上可以看出，噪声是随机的；但在较大尺度下，噪声变得很规则，因为此时心率信号变成了噪声！

图 10.31 显示了不同尺度下相应信号波形的例子。

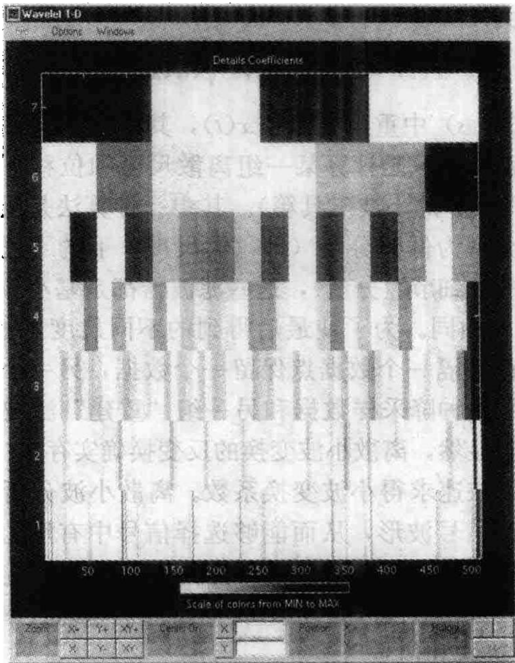


图 10.30　利用 MATLAB 工具箱作出的小波变换系数二维图形

注：所用基小波为 Daubechies 小波，分析的是图 10.2 所示的主动脉血压数据。

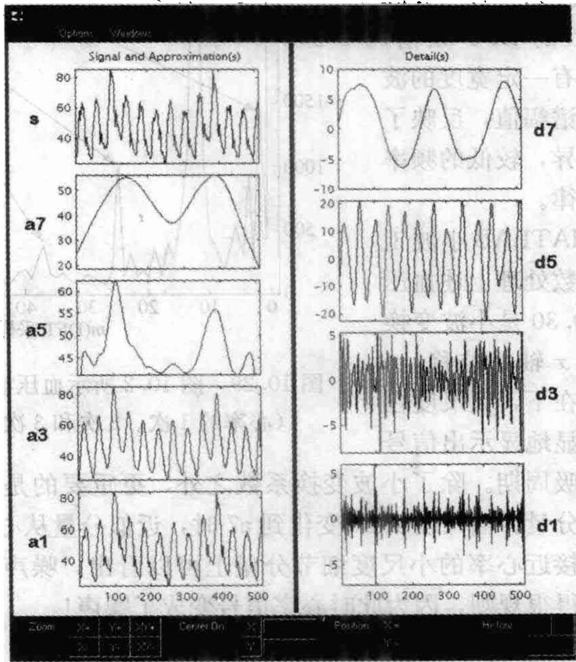


图 10.31　利用 MATLAB 工具箱生成的小波变换的几个细节分量和近似分量波形

10.9　人工智能技术

人工智能（Articial Intelligence，AI）是应用计算机系统实现智能的一个广阔研究领域。AI 系统有几种不同的实现方法，包括模糊逻辑、神经网络和专家系统等。AI 的主要目标是制造智能机器，用于不利的或者不可预测的工作环境。"智能"这个名称用于机器时，是指基于计算机的系统，它们可以与环境产生相互作用，并能够适应环境的变化。而适应算法的实现是基于对环境的定性和定量模拟。换句话说，AI 技术的基本目标是制造出更加具备人类推理能力、决策能力和适应能力的机器。

机器智商（Machine Intelligence Quotient，MIQ）是机器智能水平的一个衡量指标。机器智商越高，机器自动推理和决策的能力就越高。这几年各种各样机器的 MIQ 都显著提高。很多基于计算机的消费品、工业机器以及生物医学仪器等系统都应用了较尖端的人工智能技术。模糊逻辑、神经网络等软计算技术的发展为各种机器 MIQ 的提高做出了显著的贡献。

软计算（soft computing）并不是指某种单一的算法，而是多种算法相互结合形成的综合性方法。这些算法包括模糊逻辑、神经网络、概率推理和遗传算法等。各种软计算方法可以联合使用，形成高级智能系统。由于机器必须评估的许多参数没有精确的数值，因此，软计算允许不精确。生物系统的各种参数尤其难以精确测量和计算。

10.9.1　模糊逻辑

模糊逻辑的计算基于词语而不是数字，词语描述往往不如数字精确。传统计算都用精确的数值，而人类推理一般都用词语。模糊逻辑就是用语言变量来模拟人类的推理过程。语言变量（linguistic variables）就是用于描述参量的词语。例如，用语言变量描述体温时，可以有"高烧"、"高于正常值"、"正常"、"低于正常值"、"冰冷"等词。这些语言变量比 105.0、98.9、98.6、97.0、27.5 等华氏温度数值要模糊。

经典数学定义的数字集合称为分明集，而模糊系统的基本元素被称为模糊集。A =[0，20] 是分明集的一个例子。分明集具有精确定义的数值边界。而模糊集则没有明确的边界。例如根据人的年龄进行分类，如果用分明集，不同的年龄组可以分为 A=[0,20]、B=[30,50]和 C=[60,80] 等。图 10.32a 显示了 A、B 和 C 集合的特征函数。该函数值要么是 0，要么是 1，取决于某人的年龄是否在集合 A、B 或 C 的界限之内。分明集缺乏灵活性，如果某人 25 岁，或者 57 岁，那么这个人就不能被归类了。

如果年龄组用模糊集来表示，那么各组之间就不存在精确划分，而是用"青年"、"中年"、"老年"等语言变量来分类。图 10.32b 显示了年龄分类的模糊集。注意各个类别之间有重叠。词语是大概描述而不是精确描述。例如，30 岁女人看上去

比 6 岁男孩要老，但比 80 岁老人要年轻得多。对于模糊集，数值 1 表示 100%隶属于某个集合；数值 0 表示不属于某个集合；0 与 1 之间的所有数值表示隶属某个组的不同程度。例如，35 岁的人，50%隶属于“青年”集，50%隶属于“中年”集。

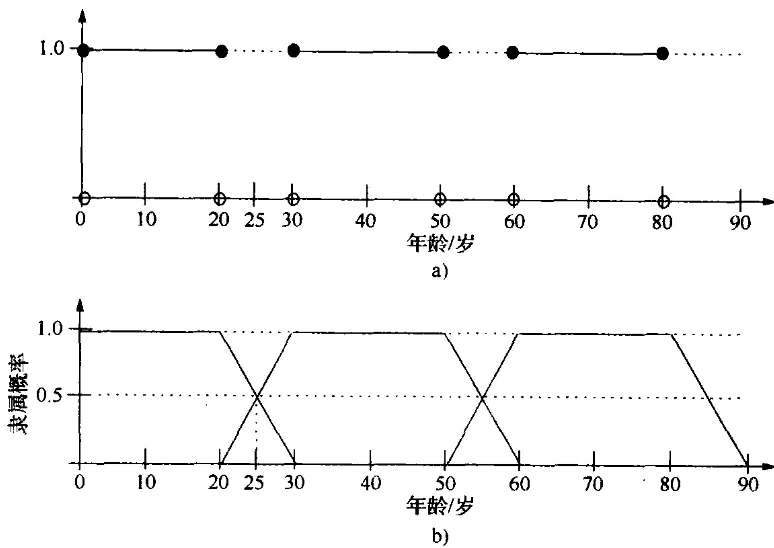


图 10.32 根据人年龄分类的分明集（图 a）和模糊集（图 b）

与经典数学的分明集一样，模糊集上也定义了运算。图 10.33a 所示是模糊集的交集运算，图 10.33b 是其并集运算，而图 10.33c 则是其补集运算，图中的实线表示运算结果。

虽然上述按年龄分类的简单例子很容易构建模糊集，但是对于比较复杂的应用，模糊集的建立也会很复杂。语言变量可以用数学式表示并用计算机处理。模糊集一旦建立，相应的规则也就形成了。模糊逻辑是基于规则的逻辑。模糊系统由很多条规则组成，并且，一般的规则都是语言变量组成的“if/then”语句。例如，按年龄分类的模糊集的两条简单规则是

if 病人是“青年”，then 用“治疗方案 A”。

if 病人是“中年”或“老年”，then 用“治疗方案 B”。

根据变量归于某个组别的隶属度来确定使用哪条规则并采取哪类措施，如采用哪种治疗方案。去模糊法用于确定选用哪条规则来产生模糊系统的最终输出。

在很多应用情况下，模糊逻辑比传统的数值计算有明显的优势。如果信息量很有限或者很复杂，不能进行精确的数值计算，那么模糊逻辑就特别有用，因为它允许不精确。如果不能建立精确的数学模型，模糊逻辑也有应用价值。但是，如果某个系统可以用数学模型精确描述，那么模糊逻辑方法通常不会比传统方法好。

在生物医学应用的各种信号分析和计算中，常会遇到某些信号的特性用传统计算方法较难处理，但很适合用模糊逻辑来处理。因为许多生物现象还没有完全弄清

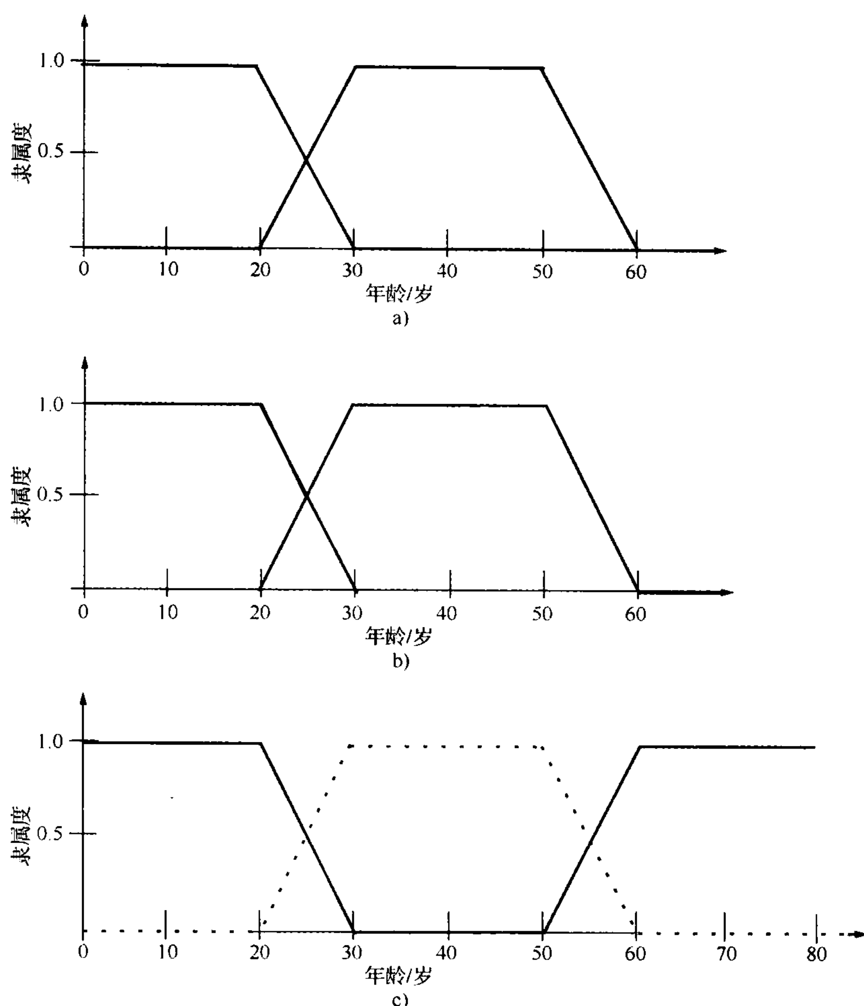


图 10.33　模糊集的交集运算（图 a)：“青年”与“中年”；并集运算（图 b)：“青年”或“中年”；补集运算（图 c)：非“中年”

楚，并且非常复杂，各个生物系统之间的差别也非常大。另外，现有的传感器等生物检测仪器具有局限性，很多生物系统的关键参量还不能精确测量。例如，传感器只能断断续续地测量某些生物量，或者测量时包含了其他参量的影响而不能只单独测量某个参量。如血糖传感器不仅对血糖敏感，而且对血液中的尿素等其他元素也敏感。模糊逻辑有助于补偿传感器的这些缺陷。

模糊逻辑在很多生物医学工程领域都有应用。例如，自动给药的闭环药物输送控制系统就是用模糊逻辑开发的。模糊逻辑在麻醉剂药物输送系统的给药控制中特别有用，因为用传统计算方法很难精确计算不同病人应该注射的麻醉药物剂量。另外，模糊逻辑还用于开发瘫痪病人神经修复的新方法。运动神经修复术利用模糊逻

辑系统控制在腿部肌肉上施加的电刺激，可以使瘫痪病人能够行走。

例 10.26　利用模糊系统可以根据心率的变化将病人分类，该系统有助于确定病人是否具有正常静息心率、或心动过缓、或心动过速。心动过缓是指静息心率小于 60 次/min 的心律失常，而心动过速则是指静息心率大于 100 次/min 的心律失常。正常静息心率的范围是 70～80 次/min。请问描述人体静息心率可以用哪 3 个语言变量?

解：可以用的语言变量名称有很多，关键是要选择可以较好地描述类别和所述对象的名称。例如，可以用“慢”、“正常”和“快”，或者直接用“心动过缓”、“正常”和“心动过速”。■

10.9.2　人工神经网络

人工神经网络（Artificial Neural Networks，ANN）是真实生物神经网络的理论模型。人类大脑是最高级的生物神经网络，它由高度关联的数十亿个脑细胞（即神经元）组成。如此高度关联的神经元结构具备极强的计算能力，远远超过最高级的计算机。人、哺乳动物、甚至简单的无脊椎动物（如苍蝇）的大脑都很容易根据经验学会识别相关的感官信号（如声音和图像等），并对周围环境的变化做出反应。人工神经网络的设计就是要模仿并复制出真实大脑的功能。

ANN 要比生物神经网络简单得多，较复杂的 ANN 也只包含数千个神经元和数百个连接。虽然比较简单，但是，ANN 的目标是要使计算机系统具有与真实大脑类似的学习、处理和适应能力。人工神经网络可以通过学习来识别某些输入，并针对给定输入产生特定的输出。因此，人工神经网络通常用于生物信号的模式识别和分类。

ANN 由多个互相关联的神经元组成，各种不同的神经元都可以在 ANN 中表示出来。并且各个神经元分层排列，各层神经元可以与其他神经元包括处于其他层的神经元相联系，神经元之间相互联系的方式决定了 ANN 的结构。ANN 的结构有很多种，适用于不同的应用需求。图 10.34 所示是一种简单的结构，它共有 6 个神经元，分成 3 层。其中，“输入层”是第一层，有 2 个神经元，它们接收神经网络的输入；“中间隐层”有 3 个神经元，用于处理信息；“输出层”有 1 个神经元，它是 ANN 的输出。

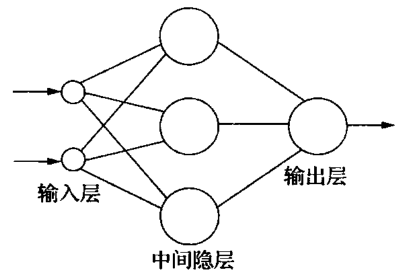


图 10.34　包含 3 层共 6 个神经元的简单人工神经网络示意图

神经元之间的关联用数学方程来描述。图 10.35 的示意图就表示了神经元及其输出的数学计算方法。计算该神经元总输入的方程是

$$x=(\text{输入 }1\times\text{权重 }1)+(\text{输入 }2\times\text{权重 }2)+\text{偏置} \tag{10.57}$$

神经元的输出由数学方程 $g(x)$ 给定，最常用的 $g(x)$ 是阈值函数和非线性 Sigmoid 函数（即 S 型函数）。利用 S 型函数的神经元输出 y 可以用以下简单公式来计算：

$$y=1/(1+e^{-x}) \tag{10.58}$$

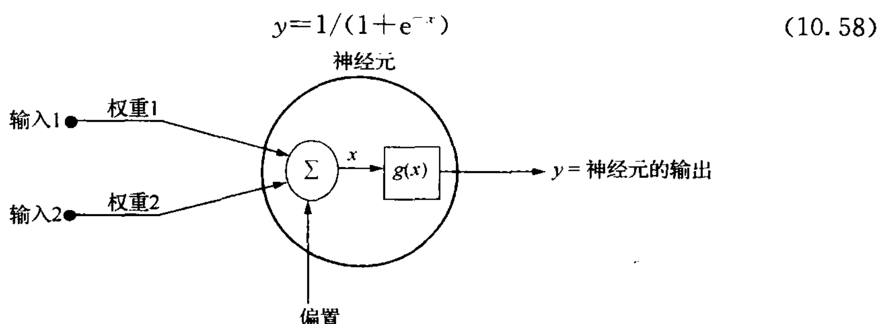


图 10.35　神经元输入和输出的数学关系

在生物信号处理的应用中，ANN 输入层接收的输入可以是原始数据，也可以是经过预处理的信号，或者是从生物信号中提取出来的某些特征量。原始数据通常是采样得到的数字化信号。预处理过的信号是指在输入神经网络之前已经经过变换、滤波等处理的信号。另外，某些特征量也可以作为神经网络的输入，例如，阈值、重复出现的特定波形、各种波形之间的时间间隔等。

ANN 必须学会识别输入信号中所包含的各种特征或者模式，但并不是一开始就会，而是必须经过训练才能学会。训练过程中，用户给 ANN 输入各种不同的重要样本事例，每个样本都需要重复输入多次。等到所有样本都试过多次之后，ANN 就具备了对特定输入产生特定输出的能力。

ANN 有各种各样学习模式，大致可以分为两大类：无监督学习和监督学习。在无监督学习中，给定训练样本集的输出是未知的，ANN 必须进行一种自组织。在这种学习过程中，ANN 要学会识别输入样本的共性，并对各种不同类型的输入产生特定的输出。生物信号处理用到的无监督学习 ANN 包括 Hopfield 神经网络和自组织特征映射网络。

在监督学习中，训练样本集的理想输出是已知的。特定输入的 ANN 输出与理想输出进行比较。理想输出作为目标，对于每个给定输入样本都计算目标与实际 ANN 输出之间的差别。监督学习的一种常用训练方法是反向传播（backpropagation，BP）算法。应用反向传播进行训练的多层感知器已经被用于生物信号处理。

反向传播算法可以使 ANN 的误差最小化。ANN 的误差就是 ANN 对于某输入所产生的实际输出与其目标输出之间的差别。反向传播用梯度法使网络误差最小化，也就是，沿着误差斜率下降的方向进行计算，逐渐逼近目标，就像球体滚下山坡一样。“反向传播”这个名称是指改变 ANN 参数以减小误差的一种方式。首先给网络中的每个神经元设定一个误差值，然后计算每个神经元的相对误差，并改变神经元之间的连接强度使误差最小化。神经元之间的连接强度就是图 10.35 中所示的权重。神经元的误差计算和权重变化从输出神经元通过 ANN 反向传播到输入神经元。反向传播是求取产生最小网络误差的最优权重数值的一种方法。

ANN 适用于各种生物信号处理，它也是一种非线性统计分析的方法，经常用于

模式识别和分类。另外，对于具有高复杂性或者包含高噪声水平的信号，已经证明ANN比传统方法执行速度快且准确。而且，ANN还可以解决没有算法解的问题，也就是不能用传统计算机编程解决的问题。由于ANN会学习，因此不需要用算法来解题。

随着人工智能技术的发展，ANN已经被广泛用于生物信号处理以及生物医学仪器。从ECG和EEG的信号分析到医学图像的识别和各种疾病的诊断都已经有ANN的应用研究。在神经病学中，ANN已被用于研究癫痫、帕金森症和老年痴呆症等大脑疾病所产生的脑部特征变化。ANN也用于心律失常ECG信号的特征提取和分类。有人曾经将ANN用于急救病人心脏病发作的诊断，研究结果表明，总体而言，ANN能够比急救室医生更好地诊断心脏病发作状况。ANN的优点是不会受到疲劳、注意力分散、情绪压力等的影响。随着人工智能技术的发展，ANN将成为许多生物信号处理的一种高级工具。

例 10.27 某神经网络中的一个神经元有3个输入，并用S型函数计算神经元的输出，设3个输入值分别是0.1、0.9和0.1。训练之后，3个输入所对应的权重分别是0.39、0.72和0.23，偏置是0.48。请问神经元的输出是多少?

解： 首先，利用式（10.57）计算输入之和，即

$$x = (\text{输入}1\times\text{权重}1) + (\text{输入}2\times\text{权重}2) + (\text{输入}3\times\text{权重}3) + \text{偏置}$$
$$= 0.1\times0.39+0.9\times0.72+0.1\times0.23+0.48$$
$$= 1.19$$

然后，利用式（10.58）计算神经元的输出，即

$$y=1/(1+e^{-x})=1/(1+e^{-1.19})=0.77$$

■

习　　题

1. 快速跑过街道时，你腿部的神经会产生哪些类型的生物信号?

2. EEG记录的是什么生物信号？请说明信号的来源和特征。

3. 请说明在自行车比赛过程中正常人体心脏ECG可能产生的变化。

4. 用16位A/D转换器采样某个生物信号的模拟量，其最小电压为−30mV，最大电压为90mV，则A/D转换的分辨率为多少?

5. 用200～250 Hz采样频率可以正确采样某块骨骼肌活动所产生的EMG信号，请问原始EMG记录信号中的最高频率是多少?

6. 有两个信号 $x_1(t)$ 和 $x_2(t)$，其幅值频谱如图10.36所示。请计算以下信号的Nyquist频率：

(a) $x_1(t)$

(b) $x_2(t)$

(c) $x(t)=x_1(t)*x_2(t)$（提示：应用卷积定理）

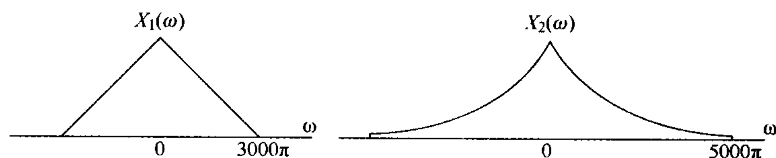


图　10.36

7. 假设信号为 $x(t)=3+\sin(2\pi100t)+\cos(2\pi250t+\pi/3)$，请计算其 Nyquisti 频率。

8. 频率为 125 kHz 的正弦波信号用 70 kHz 的频率采样，请问采样所得信号的频率是多少？

9. 脑电图 EEG 信号的最大频率是 300 Hz，用 A/D 转换器将其采样并转化为二进制数据序列。

(a) 如果信号采样频率要求比 Nyquist 频率高 50%，请计算该采样频率；

(b) 采样数据的量化要求有 2048 个水平，请问每个采样数据要用几位二进制数来表示？

10. 请写出图 10.37a 所示信号的指数傅里叶级数。

11. 请写出图 10.37b 所示信号的指数傅里叶级数。

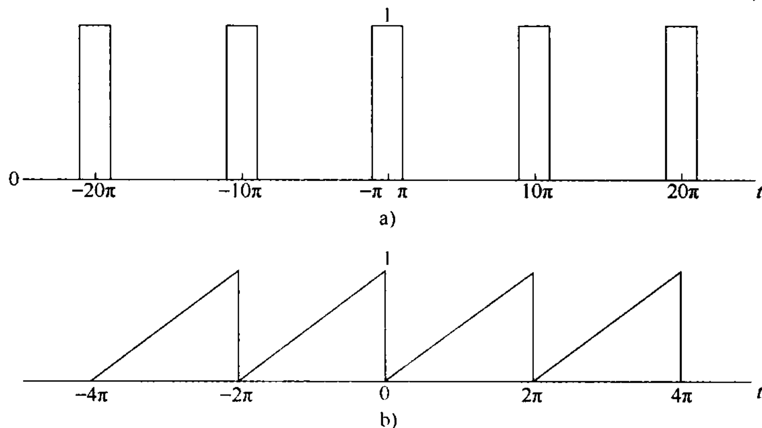


图　10.37

12. 图 10.38 所示信号 $f(t)$ 是一个周期信号，请写出其三角傅里叶级数。

13. 请写出如下三角傅里叶级数 $f(t)$ 的简化三角傅里叶级数：

$$f(t)=3+3\cos(t)+2\cos(2t)+4\sin(2t)-4\left(\frac{e^{j4t}+e^{-j4t}}{2}\right)$$

14. 请解释为什么指数傅里叶级数要用到

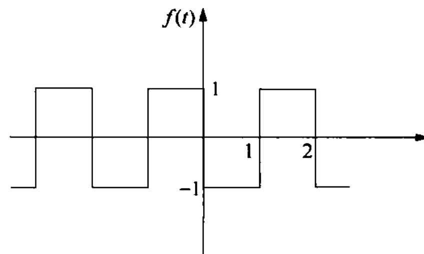


图　10.38

负频率。

15. 请写出以下函数的傅里叶变换

(a) $u(t)$

(b) $e^{-at}u(t)$

(c) $\cos(at)u(t)$

16. 请写出 $f_1(t)=e^{-3t}u(t)$ 的傅里叶变换。

17. 请写出 $f(t)=e^{-3|t|}$ 的傅里叶变换，并作出其时域和频域曲线。提示：将变量代换成某些具体数值，并计算出曲线上某些点的值，再作曲线。

18. 请证明傅里叶变换的时移性质。

19. 假设 $x(t)=e^{-at}u(t)$，且 $h(t)=e^{-bt}u(t)$，其中 a 和 b 为大于 0 的常数，请解释为什么在频域计算卷积 $x(t)*h(t)$ 更方便。

20. 用一个持续时间为 50ms，幅值为 1mA 的窄电流脉冲刺激细胞膜，假设细胞膜时间常数为 10ms，请计算细胞膜的电压输出。

21. 某细胞的离子交换过程可以用冲激响应 $h(t)=e^{-4t}u(t)$ 来估计，

(a) 如果这个 $h(t)$ 的估计是准确的，那么研究人员可以从中获得什么启发？

(b) 如果钠离子以 $x(t)=3u(t)-3u(t-2)$ 的短脉冲形式注入细胞，即脉冲持续时间为 2 s，那么，细胞将会如何反应（如泵出离子等）？请求解时域的输出响应。提示：将输入信号与冲激响应进行卷积。

22. 将心室颤动时记录的某段心电图信号进行数字化，信号的起始几个数据点的值为：−90.0、10.0、−12.0、−63.0、7.0、−22.0，单位为 mV。该生物信号数据序列的 Z 变换是什么？

23. 以下方程描述的系统中，哪些是线性的，哪些不是？

(a) $\dfrac{dy}{dt}+2y(t)=f^2(t)$

(b) $\dfrac{dy}{dt}+3ty(t)=t^2f(t)$

(c) $\dfrac{dy}{dt}+2y(t)=f(t)\dfrac{df}{dt}$

(d) $y(t)=\int_{-\infty}^{t}f(\tau)d\tau$

24. 请分析以下数字滤波器的特性

$$y(k)=\frac{1}{4}x(k)+\frac{1}{4}x(k-1)+\frac{1}{2}y(k-1)$$

写出其冲激响应、$H(z)$ 和 $H'(\Omega)$，用 MATLAB 程序计算 $|H'(\Omega)|$（其中 $0<\Omega<\pi$）并作图显示。分析该滤波器与例 10.23 所示滤波器之间的不同之处。为什么这是一个更好的低通滤波器？如果输入序列为 $x(k)=100\sin\left(\frac{\pi}{2}k+\frac{\pi}{8}\right)$，则输出是什么？如果输入序列为 $x(k)=100u(k)$，输出又是什么？

25. 请写出以下函数的 Z 变换：

(a) $x(k)=u(k)$

(b) $x(k)=a^k u(k)$

(c) $x(k)=\cos(bk)u(k)$

26. 请写出以下数据序列的 Z 变换

(a) $x(k)=(1/2)^k u(k)$

(b) $x(k)=\cos(\Omega k)u(k)$

27. 请计算以下离散系统的前 4 个输出：

$$y(k)-3y(k-1)+2y(k-2)=f(k-1)$$

假设 $y(-1)=1, y(-2)=3$，且 $f(k)=3^k u(k)$

28. 请计算以下离散系统的前 4 个输出：

$$y(k)-2y(k-1)+2y(k-2)=f(k-2)$$

假设 $y(-1)=1, y(-2)=0$，且 $f(k)=u(k)$

29. 请编写 MATLAB 程序，证明随着重复实验次数的增加，多次实验结果叠加平均之后，随机噪声会趋于零。

30. 糖尿病的正确治疗需要以准确的血糖测量作为依据。葡萄糖是人体内一种主要碳水化合物，它在体内循环，给细胞提供能源。正常人体内的胰岛素可以促进葡萄糖从血液到肌肉和脂肪组织的输送，从而调节血液中的葡萄糖浓度。糖尿病患者由于不能适当调节葡萄糖水平，血糖浓度会变得过高。请说明如何在测量血糖浓度的系统中应用模糊逻辑实现控制？与传统的系统相比，模糊逻辑系统的优点是什么？

31. 请说出 3 种可以应用人工神经网络处理的不同生物信号，并针对你所列举的应用，至少说出 2 点人工神经网络优于传统信号处理方法的优点。

32. 假设例 10.26 中的模糊集已经确定，静息心率为 95 次/min 的人归于正常集合的隶属度为 75%，归于心动过速集合的隶属度为 25%。而静息心率为 65 次/min 的人归于正常集合的隶属度为 95%，归于心动过缓集合的隶属度为 5%。请作图表示该模糊集。

推荐阅读资料

Akay, M. (1994). *Biomedical Signal Processing*. Academic, San Diego, CA.

Akay, M. (Ed.) (1998). *Time Frequency and Wavelets in Biomedical Signal Processing*. IEEE, New York.

Bauer, P., Nouak, S. and Winkler, T. (1996). A Brief Course in Fuzzy Logic and Fuzzy Control, (http://*www.flll*.uni-linz.ac.at/fuzzy). Fuzzy Logic Laboratorium Linz-Hagenberg, Linz, Austria.

Bishop, C.M. (1995) *Neural Networks for Pattern Recognition*. Oxford Univ. Press, New York.

Bruce, E.N. (2000). *Biomedical Signal Processing and Signal Modeling*. Wiley-Interscience, New York.

Ciaccio, E.J., Dunn, S.M. and Akay, M. (1993). Biosignal pattern recognition and interpre-

tation systems: Part 1 of 4: Fundamental concepts. *IEEE Eng. Med. & Biol.* **12**, 810–897.
Ciaccio, E.J., Dunn, S.M. and Akay, M. (1993). Biosignal pattern recognition and interpretation systems: Part 2 of 4: Methods for feature extraction and selection. *IEEE Eng. Med. & Biol.* **12**, 106–113.

Ciaccio, E.J., Dunn, S.M. and Akay, M. (1994). Biosignal pattern recognition and interpretation systems: Part 3 of 4: Methods of classification. *IEEE Eng. Med. & Biol.* **12**, 269–279.

Ciaccio, E.J., Dunn, S.M. and Akay, M. (1994). Biosignal pattern recognition and interpretation systems: Part 4 of 4: Review of applications. *IEEE Eng. Med. & Biol.* **13**, 269–273.

Cohen, A. (1986). *Biomedical Signal Processing: Volume I Time and Frequency Domain Analysis*. CRC, Boca Raton, FL.

Cohen, A. (1986). *Biomedical Signal Processing: Volume II Compression and Automatic Recognition*. CRC, Boca Raton, FL.

Dempster, J. (1993). *Computer Analysis of Electrophysiological Signals*. Academic, San Diego, CA.

Devasahayam, S.R. (2000). *Signals and Systems in Biomedical Engineering: Signal Processing and Physiological Systems Modeling*. Kluwer Academic, New York.

Haykin, S. (1994). *Neural Networks—A Comprehensive Foundation*. Macmillan College, New York.

Northrop, R.B. (2003). *Signals and Systems Analysis in Biomedical Engineering*. CRC, Boca Raton, FL.

Onaral, B. (Ed.) (1995). Biomedical signal analysis. In *The Biomedical Engineering Handbook* (J.D. Bronzino, Ed.). CRC, Boca Raton, FL.

Oppenheim, A.V. and Schafer, R.W. (1975). *Digital Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ.

Oppenheim, A.V., Willsky, A.S. and Young, I.T. (1983). *Signals and Systems*. Prentice-Hall, Englewood Cliffs, NJ.

Roberts R.A. and Mullis, C.T. (1987). *Digital Signal Processing*. Addison-Wesley, Reading, MA.

Smith, M. (1996). *Neural Networks for Statistical Modeling*. International Thomson Computer, Boston, MA.

Stearns, S.D. and David, R.A. (1993). *Signal Processing Algorithms in Fortran and C*. Prentice-Hall, Englewood Cliffs, NJ.

Thompkins, W.J. (1993). *Biomedical Digital Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ.

Williams, C.S. (1993). *Designing Digital Filters*. Prentice-Hall, Englewood Cliffs, NJ.

Zadeh, L.A. (1987). *Fuzzy Sets and Applications*. John Wiley, New York.

Ziemer, R.E., Tranter, W.H. and Fannin, D.R. (1993). *Signals and Systems: Continuous and Discrete*, 3rd Ed. Macmillan, New York.

第 11 章 生物电现象

John Enderle 博士[⊖]

本章目录

⊖ Joseph Bronzino 参与了本章的撰写工作。——原文注

本章主要内容：

- 生物电的发展历史；
- 神经元之间信号传导的定性描述；
- 一种或多种离子所产生的膜电压；
- 跨膜电流脉冲引起的膜电压变化；
- 膜电压随距离的变化；
- 电压钳实验技术和动作电位；
- 动作电位的 Hodgkin-Huxley 仿真模型。

11.1 绪论

本书第 3 章已经简要介绍了神经系统和神经元的基本概念，本章将深入讲解神经元静息和兴奋时的特性，并在此基础上，利用神经元通用仿真软件 GENESIS（详见参考资料中 J. M. Bower 和 D. Beeman 的著作）建立比较完整的单神经元模型和多神经元模型，并根据更精确的离子通道机制和神经网络扩展 Hodgkin-Huxley 基本模型。这些模型是了解神经系统机制及其功能的重要基础。

本章所介绍的神经元模型具有较长的发展历史，并且，如今的新发现仍然在不断更新已有的理论和模型。神经元模型的生理学意义主要在于研究神经元的信息传递和存储机制，其工程学意义则是将神经元作为计算机结构和人工神经网络设计的仿真模板。为了全面了解神经元的工作机理，首先必须应用生物物理学、生物化学以及电路理论等方法，认识细胞膜的静息特性，然后才能进一步从本质上了解通过动作电位实现的神经元信息传导和处理过程。

1952 年 Hodgkin 和 Huxley 两人发表的论文描述了一系列实验，并在此基础上建立了动作电位模型。这项工作的成功使两位作者与 John Eccles 一起获得了 1963 年诺贝尔奖，本章 11.6 节将介绍他们的工作。也许有人会问：1952 以来已经经历了那么大的发展和变化，如今还有必要在教科书中讲述那时的 Hodgkin-Huxley 模型吗？答案很简单，首先，这个模型是少见的且不会过时的经典之作，值得介绍；其次，目前具有的所有其他模型其实都源于此模型，并且将来新出现的模型可能也是如此。

本章的内容安排如下：第 11.2 节简单介绍生物电发现的历史，初次阅读本章时可以跳过这一节；第 11.3 节讲述神经元的结构；第 11.4 节介绍生物物理学和生物化学的一些基本方法，用于阐述神经元的静息特性；第 11.5 节讲解由电阻、电容以及电压源组成的静息细胞膜的等效电路模型；第 11.6 节讲述神经元的 Hodgkin-Huxley 模型，包括有关实验以及动作电位的数学模型。

11.2 生物电发现的历史

11.2.1 Galvani 与 Volta 之争

1791 年，《Proceedings of the Bologna Academy》杂志刊登了一篇论文，用实验结果证明了动物电的存在，这篇论文闻名遐迩，其作者就是 Luigi Galvani。当时，该论文引起了科学界的轰动，并且引发了一场争议，就论文中的实验结果产生了两种不同的解释，最终导致了电生理学和电气工程学两个不同学科的创立。Galvani 确信他所观察到的蛙腿肌肉收缩现象是由于动物身上发出的某种形式的电引起的；但是，当时任 Padua 大学物理学教授的 Allesandro Volta 却认为这种“电”并不是源于动物自身，而是来自实验所使用的两种不同金属。其实，这两种截然相反的观点都很重要。下面就详细介绍当时所具备的科学知识体系、这两种观点的理论基础以及这场争论的最终结果。

11.2.2 18 世纪的电学

1800 年以前，人们已经积累了大量有关普通电学和生物电的详细记载。例如，埃及人和希腊人早就知道某些鱼能够释放出很高的电压，电击水中其他生物；希腊人还发现，用猫皮摩擦树脂（即琥珀，希腊语称为“elektron”），或者用丝绸摩擦玻璃，都可以产生静电现象。小亚细亚人 Thales 在公元前 600 年就指出，用布猛烈摩擦琥珀时，琥珀就会具有某种“吸引力”，可以吸引一定距离范围内的谷壳、纸屑和丝线等小物体。当时静电被看成一种神秘莫测的事物。

与 Thales 时隔两千多年之后，英国医师 Willian Gilbert 继续了 Thales 的研究。他发现，除了琥珀之外，玻璃、玛瑙、钻石、蓝宝石以及其他很多材料经过摩擦之后都可以具有希腊人所说的那种吸引力。但是，Gilbert 没有报道小物体也可能被排斥。直到一个世纪之后，法国人 Charles DuFay（1698—1739）才发现静电排斥现象。

后来，摩擦起电有了进一步发展，人们发明了旋转摩擦机器来持续不断地产生大量静电荷。第一台摩擦发电机是德国人 Otto von Guericke（1602—1685）发明的。于是，摩擦起电在 18 世纪成为热门研究，人们通过各种实验揭示了许多新的电学特性。例如，英国人 Stephen Gray（1666—1736）证明电可以沿着用丝线悬挂起来的普通麻绳流到几百英尺之外。于是，他推测电是一种“流体”。但是，如果将丝线换成金属线来悬挂麻绳，他发现电荷很快就会消失。从此，人们又逐渐认识到某些材料可以导电，而另一些材料则是绝缘体；并且认为丝绸、玻璃、树脂之类的“电体”带有电荷，而金属和水之类的“非电体”会传导电荷。Gray 还发现，不需要直接接触，电就可以从邻近的一个充电物体传导到另一个物体上，这就是感应起电现象，根据此原理，后来很快发明了产生静电荷的机器。

法国科学院的 Charles F. DuFay 当时对 Gray 的实验很感兴趣，他用大量的测试

证明，除了金属、太软的物体或者流体之类的材料不能用来摩擦生电之外，实际上几乎所有其他材料都能够生电。但是，他后来又发现，如果把金属绝缘，那么金属可以携带的电荷其实最多。他还发现，摩擦过的玻璃会排斥小金片，而摩擦过的琥珀、树胶和蜡却会吸引小金片。他由此认为实际存在两种电“流”，并把它们分别称为“玻璃性”的和“树脂性”的电。他还指出，不同种类的电荷可以互相吸引，同种电荷则互相排斥。这一研究结果表明了两种“电”的存在。

英国美洲殖民地的 Benjamin Franklin（1706—1790）也对电很感兴趣，并通过实验提出了“单种电流”的假说。他认为只存在一种电，摩擦生电只不过使这种电产生了分离，从而一个物体带电过剩，而另一个物体则缺电。他指出，“电”是所有物体都具有的元素，并且通常处于平衡的中性状态。物体之间相互摩擦等作用会引起电荷的过量或缺失，从而导致了不平衡。因此，摩擦起电只是分离电荷，并不创造新的电荷。Franklin 的这种假说其实隐含了电荷守恒定律的思想，也就是得到的电荷与失去的电荷是相等的。他还推测，玻璃被摩擦之后会带上过剩的电荷，他称之为正电，并由此建立了一直使用至今的从正到负的电流方向。现在大家都知道产生电流的电子流动的方向其实与这种传统电流方向正好相反。

直到 18 世纪末，有关静电的知识主要还是定性描述。实验检测方法虽然不少，但都不是定量测定方法，因而一直没有建立电荷之间的数学关系。不过，这些实验为后来有关电学定律的建立打下了基础，导致了静电场力现象的定量描述。

有关静电场力的定律出现在法国，主要贡献者是物理学家兼工程师 Charles A. Coulomb（1726—1806）。Coulomb 证明，当两个带电物体相互靠近放在一起时，两者之间存在一种作用力。在这种实验现象的基础上，他进一步推导了可以完整地描述力的大小的通用公式，提出带电物体之间的吸引力或者排斥力与距离的二次方成反比关系，当时这个定律成了认识物质电荷效应的主要理论基础。至此，虽然已经有了上述这么多发现，但是，值得注意的是，在 Galvani 和 Volta 之前还没有人发现能够输送持续流动电“流”的物体，这种电“流”就是我们现在所说的电荷的电流。

Franklin 既是政治家、外交官、出版商以及美国独立宣言和宪法的起草人之一，也是一位狂热的实验家和发明家。1743 年 37 岁的他在波士顿满怀强烈兴趣看完一场静电表演之后，就下定决心要亲自研究这种奇怪的现象。于是，他立即购买并设计了各种各样的仪器装置，成了一名电学爱好者，花了好几年时间进行静电效应的实验研究。

科学家 Franklin 家喻户晓的故事可能就是他在 1752 年 6 月的一天在美国费城暴风雨中放飞风筝的实验。虽然那时很多欧洲研究人员都已经推测到电和闪电是一回事，但是，Franklin 最早用实验证明了闪电是一种巨型放电。根据金属尖端容易吸引电火花的特点，他发明了“避雷针”。垂直安装在房顶上的避雷针可以把云层中的电荷安全地引导到大地中去。这是静电学的第一个实际应用。

伦敦皇家学会（Royal Society）非常赞赏 Franklin 的工作成果。在遥远的美洲殖

民地能够做出如此卓越的成就更使得 Franklin 名声大震，他多次以政治家和实验家的身份访问欧洲，成为欧洲社交界和科学界的名人。

11.2.3 Galvani 的实验

基于上述这些电“流”的知识背景，还有很多有关电可以激活肌肉和神经的有力证据，不难理解为什么当时生物学家会开始推测 Galen 曾经提出的调节肌肉收缩和所有神经功能的“神经流”（即当时所谓的“动物灵魂”）就是一种电。身为产科医生和解剖学家的 Galvani 决不是第一位持有这种观点的人，但是，他完成的证明电就是“神经流”的实验却导致了关键性的突破。

当时普遍认为肌肉收缩可以用某种形式的动物电来解释。到了 18 世纪，人和动物的肌肉在静电装置的放电刺激下产生痉挛的现象在实验室已经习以为常。因此，触电被认为是一种肌肉刺激。为了解释肌肉为什么会收缩，人们进行了各种各样的解剖实验，研究“金属接触”与动物组织活动之间可能存在的关系。1750 年，苏黎世（Zurich）一位生理学教授 Johann Sulzer（1720—1779）报道了一个偶然发现的现象，当他把舌头放在两片不同的金属条（如锌和铜）之间时，会感觉到一种恶心的酸味。两片金属条的末端要接触在一起才能产生这种现象；如果分开，就没有这种感觉了。Sulzer 认为金属片产生的振动刺激了舌头从而产生这种味觉现象。他用其他不同的金属也得到了同样的结果。但是，Sulzer 的发现当时并没有引起人们的关注，半个世纪之后随着新发现的出现人们才开始注意到他的实验结果。

接下来的就是前面提到的 Galvani（1737—1798）的那个幸运且非常重要的发现。Galvani 是意大利城市 Bologna 一个大家族的后裔，25 岁就成了 Bologna 大学的解剖学教授。他对电及其与肌肉、神经活动之间可能的关联具有强烈的兴趣。肢解的蛙腿标本当时是一种常用的实验材料，Galvani 在实验中就是用蛙腿研究肌肉和神经活动。他和同事们观察了动物组织对于各种刺激的反应，并注意到，如果将手术刀刀尖放在新鲜制备的蛙腿标本上，只要放在旁边的摩擦起电静电仪一放电，蛙腿就会强烈抽搐。

Galvani 在他的实验记录中写到：

“我解剖了一只蛙腿标本，把它放在桌上，桌上还有一台静电仪。我一位助手的手术刀刀尖偶然搭在了蛙小腿神经上，小腿的肌肉突然猛抽了一下。另一位做实验的助手注意到，只有在静电仪放电的瞬间才有这种肌肉收缩，他对此现象很是惊奇。当时我正在忙其他事，得知此事后，我立即重复了这个实验，用手术刀刀尖接触小腿神经的一端，助手则摇动静电仪放电。每次一放电，肌肉都会抽搐。”

凭着敏锐的洞察力和训练有素的思维，Galvani 进一步设计了一系列实验来研究这种神秘的肌肉收缩现象。经过反复实验，他发现诱发肌肉收缩所需的条件是将标本放在金属盘中，同时用金属物体接触肌肉。

Galvani 当时已经知道 Franklin 用实验证明了闪电与静电仪所产生的电具有相同的性质，于是他设计了一个实验，证明利用大气中的电也可以像静电仪那样产生同样

的结果。他把蛙腿神经搭在天线上，而把蛙脚放在电的接地点上，在暴风雨中他看到蛙腿肌肉产生了与用静电仪时相同的反应。这些实验现象引发了一系列疑问、研究和争议。

Galvani 还注意到，用铜钩穿过蛙腿标本的骨髓，并将标本挂在铁架子上，在任何天气情况下，蛙腿肌肉都会偶尔产生抽搐。在操作标本时，每当他将铜钩挂到铁架子上，两种金属一接触，肌肉就会抽搐。为了确定这种收缩是否仍然与某种大气效应有关，他在实验室房间里重复了实验。结果发现，每当穿在骨髓上的铜钩与铁盘接触时，放在铁盘中的标本就会抽搐。他感到这种现象反映了新的机理，于是就设计新的实验来寻找真正的起因。他发现用玻璃代替铁盘就不会产生这种肌肉反应，而用银盘则又可以观察到这种反应。于是他把长度相同的两片异种金属条连接在一起，并弯成一个弓，用这种双金属弓的两端接触蛙腿标本时，就又看到了同样的肌肉抽搐反应。于是，他得出结论：金属接触是肌肉收缩的作用因素，而且制作金属弓的金属种类不同，肌肉收缩的强度也不同。

现在 Galvani 面临的问题是如何解释所观察到的现象。他的实验标本其实显示了两种电效应：一种来自静电仪的放电，另一种则来自两种不同金属的接触。引起肌肉活动的电，要么存在于肌肉标本内，金属只是起到了释放电的作用；要么来源于两种不同金属的接触，肌肉标本只是一个指示器而已。

Galvani 是一位解剖学家，因此他提出的是第一种解释，将实验现象归因于生物体本身肌肉和神经中存在的“动物电”。他建立了一个生理学模型，把生物体比作“莱顿瓶”(Leyden jar)⊖，体内不同的组织会产生相反的电荷。这些电荷通过神经从大脑流向肌肉，当电荷通过金属连接释放出来时就会引起肌肉抽搐。他这样写到：“思路清楚了，动物体内存在一种内在的电。我们推测肌肉收缩现象发生时有一股很细的神经流通过神经流向肌肉，就像莱顿瓶中的电流一样。”Galvani 的假说反映了他那个时代盛行的观点，就是将生物体的活动看成是体内各个部分“灵魂”流动的结果。

1791 年，Galvani 在 Bologna 科学院学报（the proceedings of the Academy of Science in Bologna）上发表了题为“静电力与肌肉运动之间的关系”（De Viribus Electricitatis In Motu Musculari）的论文，阐述了他的实验现象和推论。这一报道引起了轰动，它暗示着生命力量之谜的新发现。不论是科学界行内人士还是行外人士，不管是意大利还是欧洲其他地方，人们都为这些发现而激动。但是，没有人比 Allesandro Volta 更专心致志地跟踪 Galvani 的发现，并且以这些发现为基础获得了更伟大的发明。

11.2.4 Volta 的解释

Galvani 的研究激发了人们巨大的兴趣。不管在哪里，只要能拿到青蛙，科学家们就重复他的实验，并都取得了成功。开始时，人们毫无疑问地接受了 Galvani 关于肌肉收缩的解释，甚至连杰出的物理学家 Allesandro Volta 也同意 Galvani 的观点，他得到 Galvani 的论文之后，就立即验证了这种现象。

⊖ 莱顿瓶是一种早期的电容器，可以充放电。——译者注

Volta 是一位受人尊敬的科学家，他 24 岁就发表了第一篇科学论文“论电的吸引力”（On the Attractive Force of the Electric Fire），分析了电场力与重力的相似性。Volta 致力于物理学和数学研究，也忙于实验研究，他具有杰出的才能，不到 30 岁就已经成为 Como 皇家学院（Royal School of Como）的物理学教授，并在那里完成了他的第一项重大科学创造，发明了起电盘（electrophorus），其实是一种电的“搬运”工具，是第一台用感应方式而不是摩擦方式可以反复产生电荷的装置。

1782 年，Volta 被聘为 Padua 大学的物理学教授。在那里他做出了第二个发明，就是电容起电盘，它是一台灵敏度很高的检测电荷的仪器。在此之前人们用验电器检测电荷的存在。验电器由一根绝缘金属棒以及悬挂在末端的一对丝线、或者一对小木髓球、或者两片金箔组成。如果金属棒的一端遇到电荷，那么，悬挂在另一端的这对小物体就会因排斥力而相互分开。分开的距离可以反映电荷的强度，因此，它其实是库伦定律的应用。

Volta 将这种验电器与他发明的起电盘结合起来，组成了一种用于科学研究的小量电荷检测器，就是电容起电盘。他还进一步将电容起电盘作为机械天平的一部分，通过重力作用来测量电荷所产生的力，这种仪器称为静电计（electrometer）。在 Volta 后来进行异种金属连接产生电的研究中，静电计发挥了重要的作用。

Volta 看到 Galvani 1791 年在 Bologna 科学院学报上发表的“静电力与肌肉运动之间的关系”一文之后，立刻产生了浓厚的兴趣。他很快重复了 Galvani 的实验，并肯定了 Galvani 关于“动物电”引起肌肉反应的结论。与 Galvani 一样，他将肌肉活动归因于肌肉和神经之间电的失平衡，金属接触之后，平衡又重新恢复。不过，在继续深人研究过程中，Volta 开始怀疑“动物电”观点的正确性。他发现了电荷平衡学说的破绽。实验中，只有通过金属连接使神经处于电路回路中时，肌肉才会收缩。

为了研究肌肉活动的真正起因，Volta 重新捡起了 Sulzer 曾经做过的实验。当他把一片锡箔放在舌尖上面，把一块银币放在舌尖下面，并用铜线将两者连起来时，他尝到了一股酸味。用银汤匙替换银币，省去铜线，直接用汤匙的柄接触锡箔，也会产生同样的效果。而当他用不同金属在舌头和前额之间形成连接时，他感觉到了一种亮光。基于这些结果，Volta 得出结论，认为他体验到的这些感觉不可能是由金属传导过来的，而一定是不同金属自身能够产生电。

经过两年的实验研究，1792 年 Volta 发表了他的结论，在赞扬了 Galvani 的惊人新发现之后，他对于这种效应的起因提出了不同的观点。1794 年，Volta 与 Galvani 彻底决裂，他直言不讳地反对“动物电”理论，提出了“金属电”理论。Galvani 本性谦和，他并不与 Volta 直接对抗，而是埋头继续做动物实验。

Volta 的论述指出 Galvani 的实验并没有表明动物电的存在，这对 Galvani 是一个巨大的打击。不过，Galvani 一直坚持动物电的观点，并且设计了第三个实验，彻底证明了生物电的存在。在这个实验中，他将青蛙一条腿的神经—肌肉标本扭转过来，使其脊柱和坐骨神经可以接触到另一条腿的肌肉。每当脊柱和坐骨神经接触大腿时，肌肉就会猛烈收缩。根据多数历史学家的说法，当时是支持他的侄子 Giovanni Aldini

(1762—1834) 说出了这个重要的实验，他侄子可能就是实验参与者之一。这个实验确凿地证明没有金属导体也可以诱发肌肉收缩。根据 Fulton 和 Cushing 的论著，Aldini 是这样写的：

“事实上，某些哲学家一直在设计不用金属就能诱发青蛙肌肉收缩的方法，我叔叔 Galvani 就想出了一种巧妙的方法，引起了我的注意，使我感到这个实验重要性，因此我很早就开始有愿望，要研究这个实验过程。早在 1794 年，我就在 Bologna 科学院公开演示过不借助金属，就可以诱发蛙腿肌肉收缩的实验。实验是这样的：我将蛙腿标本浸入浓氯化钠溶液中，然后取出，用手捏住一端，使另一端悬空，自然下垂；再用玻璃小棒挑出神经，使神经不接触肌肉，然后突然撤出玻璃棒；每当脊髓神经碰到肌肉时，肌肉就会收缩。某些说法，如刺激来自于盐溶液的作用、或者来自于神经坠落时产生的撞击等，都很容易被排除。只要在别的没有处于这种 Galvani 回路的青蛙标本上，用同样的神经去触碰肌肉，就可以得到证实，因为在这种情况下，盐溶液或者更强烈的碰撞都不会引起肌肉活动。”

Volta 和 Galvani 都拥有支持者和反对者，科学家们的观点在两种推论之间摇摆不定。这个论题很复杂，当时还没有被完全解决，不过，眼看真相大白已经为期不远。Carlo Matteucci 就在此时作出了一个伟大贡献，他肯定了 Galvani 的第三个实验，同时又有了新发现，证明骨骼肌收缩之前有动作电位的出现。Galvani 的第三个实验证实了损伤电位的存在，与此一致，Matteucci 写到：

“我将活体动物的肌肉切开一个伤口，然后往伤口深处插入蛙腿神经，神经用玻璃管挑住以绝缘。当神经纤维插入伤口内部时，蛙腿立刻强烈收缩起来。如果将神经纤维的一端接触伤口深部，另一端接触伤口边缘完好处，那么，这种收缩就可以一直持续下去。”

利用电流计，Matteucci 发现肌肉强直收缩期间，损伤处与非损伤处之间的电位差减小了。这个现象几乎吸引了所有电生理学家的注意力。此外，Matteucci 还完成了另一项重大发现，那就是伴随着骨骼肌收缩所产生的瞬间生物电变化，就是现在所谓的动作电位。他指出，收缩中的肌肉可以刺激神经，反过来，神经又引起其所支配肌肉的收缩。至此，Galvani 和 Matteucci 的实验已经完全确定了生物电的存在。不久，人们也在心肌和神经中都发现了动作电位。

另一方面，Volta 坚持电只是来源于不同金属的接触这个观点，他认为动物组织只不过起了指示器的作用。他用不同的双金属片配对进行了大量实验，得到了不同的实验结果，例如，他发现，有些金属配对可以增强肌肉的反应。

为了获得比较精确的定量测定数据，Volta 不再用肌肉和神经作指示器，而是改用他的“电容验电器”进行测量。幸亏他用了这种高级仪器，否则，由于不同金属之间的接触电位非常小，一般的箔金片验电器根本检测不出来。Volta 的电容验电器有一个固定圆盘和一个活动圆盘，两者之间用绝缘清漆涂层隔开。清漆涂层很薄，因此这种结构就形成了一个大容量电容器，可以聚集电荷。电容充电后活动圆盘被抬起，释放的电荷会使验电器的金箔片产生偏转。

Volta 系统地测试了各种不同金属的连接结果，用不同的金属做圆盘，通过电容验电器金箔片偏转的角度来测量各种金属盘组合所产生的电荷量。他还用摩擦过的玻璃棒或者树脂棒靠近验电器，来判别电荷的正负，从这些棒对于金箔片偏转作用可以看出电荷的极性。

Volta 的实验使他逐步形成了电场力和电位的概念，他推测这种力存在于不同金属的连接之中。在不断用不同的组合进行试验时，他还发现，金属和某些液体接触时也会产生电位。于是，Volta 就向导电系统中添加了盐水和稀释酸溶液等液体，并且将金属连接称为"第一类带电体"，而将这些液体称为"第二类带电体"。

Volta 发现在纯粹由不同金属组成的电路中只存在瞬间即逝的电位，而在两片不同金属片之间夹一层浸过盐水或者酸溶液的隔离层之后，电位指示就很稳定了。其实，Volta 当时是做了一个简单的电池——两种不同的金属加上液体隔离层，这是电池的最基本结构。另外，他还发现增加这种金属和溶液组合的数目，可以放大总体的电效应。于是，他就把金属盘和浸湿的隔离层一个接一个地垂直叠起来，构成了一个"电堆"，这就是历史上的第一个电池，也是 Volta 一生中最重要的发明。

11.2.5　最终结果

过了相当长的时间之后，人们才对 Galvani 和 Volta 所做的工作有了正确的认识。显然，他们两人都证明了电位的存在，只是不清楚电位的起因。Galvani 当时认为动物体内中存在着一种生理电容，这种生理电容由神经（内导体）和肌肉表面（外导体）组成，其中存储的动物电的释放可以触发肌肉收缩。现在，大家都知道是动作电位刺激引起了肌肉收缩。

有趣的是，神经系统的基本单元——神经元，即使在静息时其细胞内外之间也具有电位差。这种细胞膜的静息电位不断地受到各种输入信号的影响，当膜电位的变化达到某个阈值时，神经元就产生动作电位并沿着轴突传向远端。这是神经系统信息传导的基本机制。

Volta 发明的电池为科学界提供了第一个稳定的电压源，将它接入由导电金属或者导电液体组成的电路中，就会产生电流。这种装置的发明开创了电气工程学科。

11.3　神经元

根据合理的估计，人脑大约包含有 10^{12} 个神经元，可以分成大约 1000 种不同的类型，它们所呈现的形态都比较一致。必须注意，大脑中其实存在两大类细胞：神经元和神经胶质细胞。神经胶质细胞的数量是神经元数量的 10～50 倍，但是，由于神经胶质细胞不参与信号的传递，主要起支持神经元的作用，因此，本章关注的重点是神经元。神经元之间以及神经元与外周之间互相连接的特性决定了大脑复杂的功能，神经元的个体差异并不起作用。

如图 11.1 所示是一个典型的神经元，它由 4 个主要部分组成：胞体、树突、轴

突和突触前末梢。神经元胞体与其他细胞类似，含有细胞核以及其他给细胞提供营养的各种细胞器。与其他细胞不同的是，神经元胞体上连接了很多称为树突的短分支和一条称为轴突的长分支。树突是神经元的信号接收部位，它没有信号放大功能，只是被动地接收来自成千上万个其他神经元的信号。树突和胞体上都分布有突触连接，接受来自邻近神经元突触前末梢的信号。每个神经元一般有 10^4～10^5个突触。正如第 3 章所述，神经元之间的信息传递通过神经递质完成，神经递质可以改变突触后膜的膜电位。除了树突之外，与胞体相连的还有一条轴突，它将胞体和突触前末梢连接起来。其长度从 1 米（如人体脊髓）到几毫米（如大脑内部神经）不等，其直径变化也很大，有的不到 1μm，有的却有 500μm。一般，轴突直径越大，信号传导速度越快。轴突的信号传递速度范围为 0.5 ～120m/s。轴突的功能就是作为传输线在神经元之间或者神经元与肌细胞之间高速传送信号。较粗的轴突由一层称为髓鞘（myelin sheath）的绝缘物质包裹，髓鞘上有规律地分布着间隙，称为郎飞氏节（nodes of Ranvier）。动作电位沿着轴突传播时，从一个节跳到下一个节，就像幅值不会衰减的脉冲信号在传播。本章下面的大部分篇幅将用于详细介绍这种传播过程。轴突末端是一个网络，其分支可达 10000 个，各个分支末端被称为突触前末梢。图 3.28 显示了突触前末梢的结构。沿轴突传导过来的所有动作电位都通过每个分支到达突触前末梢。这些末梢是神经元的发送单元，它们在动作电位的刺激下会释放神经递质，神经递质经过大约 20nm 的突触间隙扩散到邻近细胞的突触后膜，并与突触后膜产生作用，使突触后膜的电位产生变化。

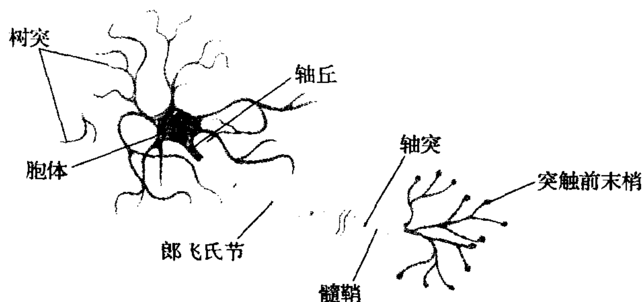


图 11.1 神经元示意图

11.3.1 膜电位

与体内其他细胞一样，神经元细胞膜内外也存在分离的电荷。如图 11.2 所示，膜外侧带正电荷，膜内侧带负电荷。膜对于离子的选择性通透形成了这种电荷分离，并由此产生了膜电位。神经元跨膜电位差大约为 60～90mV，与细胞种类有关。通常将胞外电位定义为 0mV，也就是地电位，于是，静息膜电位就是 $V_m = v_i - v_o = -60\text{mV}$。由于任何神经信号都体现在这种跨膜电位的变化之中，动作电位等信号就是膜电位变化的结果，因此这种细胞膜内外的电位差特别重要。根据定义，如果膜电位变得比静息电位更低（如－60mV 变成了－70mV），则称为超极化；如果膜电位比静息电位高（如－60mV 变成了－50mV），则称为去极化。

要建立－60mV 的膜电位，膜两侧并不需要很多分离的正、负电荷。根据电容公

式 $Cdv=dq$，或者 $C\Delta v=\Delta q$ 可以计算出所需的电荷量，其中 Δq＝电荷数×电子电荷量（即 1.6022×10^{-19}C）。如果电容 $C=1\mu F/cm^2$，并且膜电位 $\Delta v=60\times10^{-3}$V，那么，所需的电荷数约为 3.8×10^{11}个/cm²㊀。这些电荷位于膜两侧 1μm 范围之内。

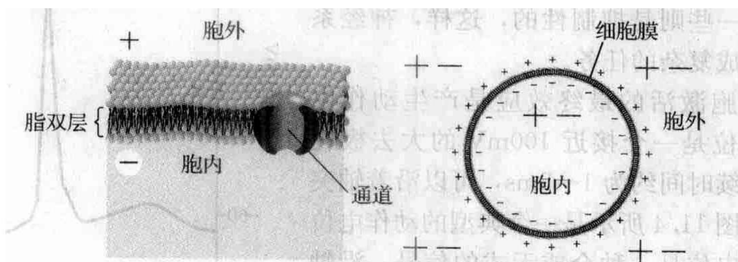


图 11.2　细胞膜两侧电荷分布示意图

注：左图表示细胞膜上正离子分布于外表面，负离子分布于内表面。右图进一步表明只有紧贴细胞膜内外表面的正负离子才对膜电位的形成产生作用，其他地方的正负离子几乎均匀分布（图中用大符号“+−”表示）。总之，细胞内侧存在净负离子，而紧靠细胞膜外侧存在净正离子。为了简化，右图的细胞膜画成圆形并忽略了轴突和树突。

分级应答和动作电位

利用突触所释放的神经递质，神经元可以改变与其相连的其他神经元的膜电位。如图 11.3 所示，神经递质穿过突触间隙，与相邻细胞的树突或者胞体上的突触后膜上的受体分子发生作用，从而改变突触后神经元的膜电位。

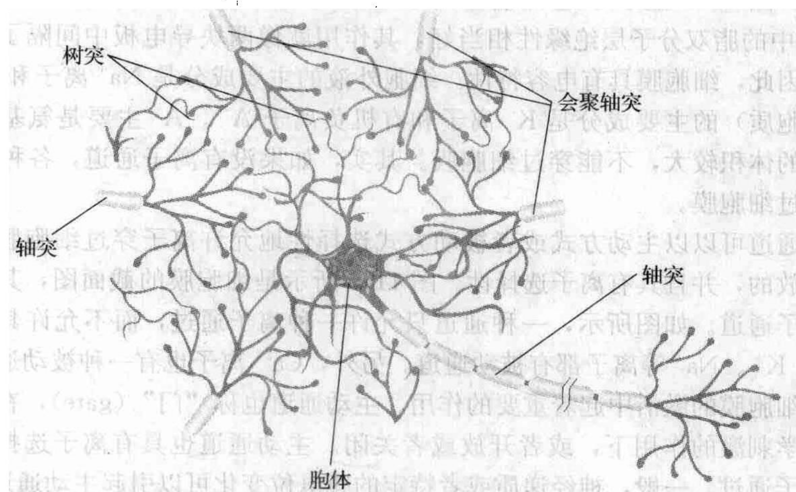


图 11.3　邻近神经元的突触前末梢在神经元树突上形成突触连接的示意图

神经递质的化学能转化为电能从而引起突触后膜膜电位的变化，该变化的大小取决于突触后膜所接受的神经递质的量以及膜去极化或者超极化的程度，因此，这

㊀ 原文此处误为 1×10^8个/cm²。——译者注

种电位变化通常称为分级应答。膜电位的分级应答也可以看做是突触上神经递质聚集和积累的结果。注意，神经元发出的信号不是兴奋性的就是抑制性的，有些突触是兴奋性的，而另外一些则是抑制性的，这样，神经系统就能够完成复杂的任务。

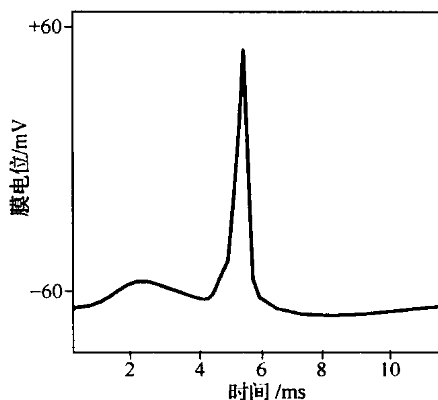


图 11.4 动作电位波形

神经细胞激活的最终效应是产生动作电位。动作电位是一个接近 100mV 的大去极化电位，其持续时间约为 1～5ms，可以沿着轴突传播出去。图 11.4 所示是一个典型的动作电位波形。动作电位是一种全或无式的信号，沿轴突传播时幅值不会衰减。当信号到达轴突末端的突触前末梢时，电位的变化引起一群神经递质小泡的释放。这是一种非常有效的长距离信号传输方法。本章下面在介绍一些基础知识之后，将详细讲解动作电位。

11.3.2 静息电位、离子浓度和离子通道

神经细胞膜上存在的静息膜电位是由于离子在细胞膜内外分布不均产生的。细胞通过选择性渗透膜，以及在后面会介绍的主动离子泵，来维持这些离子浓度的不均匀分布。图 11.2 显示了这种包含离子通道的选择性渗透膜。神经细胞膜厚约 10nm，其中的脂双分子层绝缘性相当好，其作用就像两块导电板中间隔了一层绝缘体一样，因此，细胞膜具有电容特性。细胞外液的主要成分是 Na^+ 离子和 Cl^- 离子，内液（细胞质）的主要成分是 K^+ 离子和有机负离子 A^-。A^- 主要是氨基酸和蛋白质，它们的体积较大，不能穿过细胞膜。其实，如果没有离子通道，各种离子几乎都不能穿过细胞膜。

离子通道可以以主动方式或者被动方式选择性地允许离子穿过细胞膜。被动通道总是开放的，并且具有离子选择性。图 11.5 所示是细胞膜的截面图，其中只包含了被动离子通道。如图所示，一种通道只允许一种离子通过，而不允许其他离子通过。Cl^-、K^+、Na^+ 等离子都有被动通道。另外，Ca^{2+} 离子也有一种被动通道，它在突触部位细胞膜的激活中起着重要的作用。主动通道也称“门”（gate），在外部电刺激或者化学刺激的作用下，或者开放或者关闭。主动通道也具有离子选择性，只允许某种离子通过。一般，神经递质或者特定的膜电位变化可以引起主动通道的开放。

图 11.6 所示是包含主动通道的细胞膜示意图。其中，K^+ 通道是主动通道，而 Cl^- 通道是被动通道。我们将会看到，被动通道是静息膜电位形成的机制，而主动通道是分级应答和动作电位形成的机制。

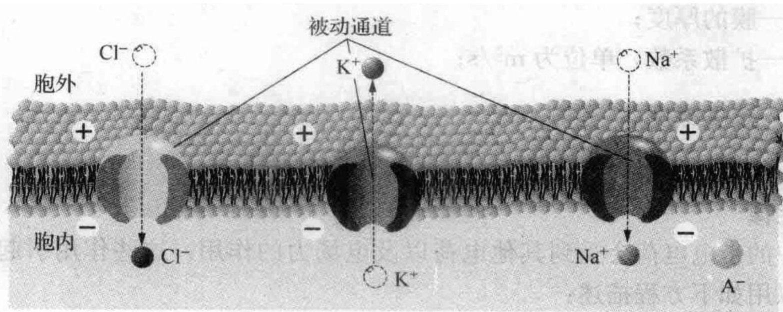


图 11.5　包含离子通道的选择性渗透膜的理想化截面图

注：膜的厚度以及通道的大小没有按照实际比例画。根据实际比例，细胞膜的厚度大约是离子大小的 20 倍、通道大小的 10 倍，而各个通道之间的距离大约是细胞膜厚度的 10 倍。注意，图中用“+”、“−”符号表示膜内外存在的电位差。通过离子特异性通道，膜对离子具有选择性通透；即图中所示的每个通道都只允许一种离子通过。

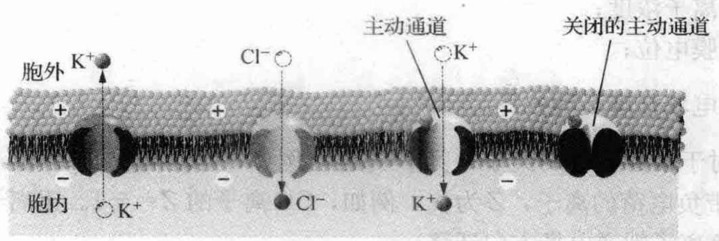


图 11.6　被动通道和主动通道是离子跨膜传输的机制

注：每种通道都是离子特异性的。图中右边的两个主动通道中左边一个处于开放状态，且只允许 K^+ 通过，而右边一个则处于关闭状态，不能通过任何离子。另外，图中还画出了一个被动 Cl^- 通道。

11.4　生物物理学基础及其相关方程

11.4.1　基本定律

膜静息电位可以用两个物理学基本定律和一个方程来描述，它们定量描述了离子梯度和电场的作用。

Fick 定律

由扩散引起的粒子流动总是沿着浓度梯度从高浓度向低浓度进行。对于细胞膜而言，扩散引起的离子的跨膜流动可以用如下方程描述：

$$J(\text{扩散}) = -D\frac{\mathrm{d}[I]}{\mathrm{d}x} \tag{11.1}$$

式中　J——扩散引起的离子流量；

$[I]$——离子浓度；

dx——膜的厚度；

D——扩散系数，单位为 m^2/s；

$\frac{d[I]}{dx}$——浓度梯度。

方程中的负号表示离子流动方向是从高浓度到低浓度。

欧姆定律

溶液中的带电电荷会受到其他电荷以及电场力的作用，这些作用引起的跨膜离子迁移可以用如下方程描述：

$$J(\text{迁移}) = -\mu Z[I]\frac{dv}{dx} \tag{11.2}$$

式中 J——在电场 $\vec{E}$ 作用下离子的迁移通量；

μ——迁移率，单位为 m^2/sV；

Z——离子价；

$[I]$——离子浓度；

v——膜电位；

$\frac{dv}{dx}$——电场强度，即 $-\vec{E}$。

注意，对于带正电荷的离子，其 Z 为正，例如，Na^+ 离子的 $Z=1$，Ca^{2+} 离子的 $Z=2$；对于带负电荷的离子，Z 为负，例如，Cl^- 离子的 $Z=-1$。正离子顺着电场方向迁移，而负离子则逆电场方向迁移。

图 11.7 所示是只对 K^+ 通透的细胞膜，图中显示了作用于 K^+ 离子的各种作用力。假设神经元细胞膜内的 K^+ 浓度高于膜外，并且膜静息电位为外正内负，那么，显然只有 K^+ 离子可以穿过细胞膜，而 Na^+、Cl^- 和 A^- 等离子由于膜上不存在对它们通透的通道，因此都不能穿过细胞膜。在这种情况下，根据 K^+ 离子浓度和膜电位的具体数值，当迁移力和扩散力达到平衡时，K^+ 的通透才会结束。也就是，随着细胞膜内的 K^+ 离子不断流向膜外，扩散引起的由内向外的化学力将减小，同时由于 K^+ 离子在胞外的聚集，由外向内的电场力不断增加，直到这两种力达到平衡时，K^+ 离子的净通透也就为 0。

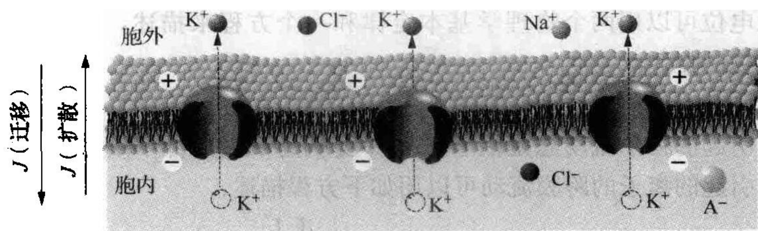


图 11.7 细胞膜上 K^+ 离子迁移流动和扩散流动方向示意图
（图中细胞膜只对 K^+ 具有通透性）

爱因斯坦方程

在渗透压作用下带电粒子在电场中的运动特性，也就是扩散和迁移之间的关系，可以用以下方程描述：

$$D=\frac{KT\mu}{q} \tag{11.3}$$

式中 D——扩散系数；

μ——迁移率；

K——Boltzmann 气体常数；

T——热力学温度，即热力学温度，单位是 K；

q——单位电荷，即 1.60186×10^{-19}C。

11.4.2 离子通透膜的静息电位

浓度梯度作用下的离子流动受到神经细胞膜的选择性通透以及由此形成的电场的影响。如前所述，离子只能穿过对它具有选择性通透的通道。为了简化，下面先考察只对一种离子具有通透性的细胞膜，然后再来考察对多种离子具有通透性的细胞膜。有趣的是，神经胶质细胞只对 K^+ 离子通透，而神经元对 K^+、Na^+ 和 Cl^- 离子都有通透。我们将会看到，如果细胞膜像神经胶质细胞那样，只对 K^+ 离子通透，那么，就可以维持正常的离子浓度梯度。

假设图 11.7 所示的细胞膜只对 K^+ 通透，并且胞内 K^+ 浓度高于胞外，那么，浓度梯度作用下的 K^+ 离子扩散方向是由内向外，即

$$J_K(\text{扩散})=-D\frac{\mathrm{d}[K^+]}{\mathrm{d}x} \tag{11.4}$$

而电场作用下的 K^+ 离子迁移方向则是由外向内，即

$$J_K(\text{迁移})=-\mu Z[K^+]\frac{\mathrm{d}v}{\mathrm{d}x} \tag{11.5}$$

这两种离子流动形成的净流量为

$$J_K=J_K(\text{扩散})+J_K(\text{迁移})=-D\frac{\mathrm{d}[K^+]}{\mathrm{d}x}-\mu Z[K^+]\frac{\mathrm{d}v}{\mathrm{d}x} \tag{11.6}$$

代入爱因斯坦方程 $D=\frac{KT\mu}{q}$，则上式变为

$$J_K=-\frac{KT}{q}\mu\frac{\mathrm{d}[K^+]}{\mathrm{d}x}-\mu Z[K^+]\frac{\mathrm{d}v}{\mathrm{d}x} \tag{11.7}$$

根据此式可以计算任意时刻、任意给定起始条件下的 K^+ 离子流量。特别是在稳态下，也就是流进胞内的 K^+ 与流出胞外的 K^+ 达到平衡时，$J_K=0$，式（11.7）可以简化为

$$0=-\frac{KT}{q}\mu\frac{\mathrm{d}[K^+]}{\mathrm{d}x}-\mu Z[K^+]\frac{\mathrm{d}v}{\mathrm{d}x} \tag{11.8}$$

由于 $Z=+1$，该式可以简化为

$$\mathrm{d}v = -\frac{KT}{q[\mathrm{K}^+]}\mathrm{d}[\mathrm{K}^+] \tag{11.9}$$

上式求积分，可得

$$\int_{v_o}^{v_i}\mathrm{d}v = -\frac{KT}{q}\int_{[\mathrm{K}^+]_o}^{[\mathrm{K}^+]_i}\frac{\mathrm{d}[\mathrm{K}^+]}{[\mathrm{K}^+]} \tag{11.10}$$

式中　　v_i、v_o——膜内、外的电压；

$[\mathrm{K}^+]_i$、$[\mathrm{K}^+]_o$——膜内、外 K^+ 离子浓度。

于是

$$v_i - v_o = -\frac{KT}{q}\ln\frac{[\mathrm{K}^+]_i}{[\mathrm{K}^+]_o} = \frac{KT}{q}\ln\frac{[\mathrm{K}^+]_o}{[\mathrm{K}^+]_i} \tag{11.11}$$

这就是以德国物理化学家 Walter Nernst 命名的 Nernst 方程，其中 $E_\mathrm{K} = v_i - v_o$ 是 K^+ 离子的 Nernst 电位。室温下，$\frac{KT}{q} = 26\mathrm{mV}$，因此 K^+ 离子的 Nernst 方程变为

$$E_\mathrm{K} = v_i - v_o = 26\ \ln\frac{[\mathrm{K}^+]_o}{[\mathrm{K}^+]_i}\mathrm{mV} \tag{11.12}$$

虽然该方程是针对 K^+ 离子导出的，但是，很容易由此推导出其他可通透离子的方程。例如，室温下，Na^+ 离子的 Nernst 电位为

$$E_\mathrm{Na} = v_i - v_o = 26\ \ln\frac{[\mathrm{Na}^+]_o}{[\mathrm{Na}^+]_i}\mathrm{mV} \tag{11.13}$$

Cl^- 离子的 Nernst 电位则为

$$E_\mathrm{Cl} = v_i - v_o = -26\ \ln\frac{[\mathrm{Cl}^-]_o}{[\mathrm{Cl}^-]_i} = 26\ \ln\frac{[\mathrm{Cl}^-]_i}{[\mathrm{Cl}^-]_o}\mathrm{mV} \tag{11.14}$$

此方程中的负号是由于 Cl^- 的 $Z=-1$。

11.4.3　Donnan 平衡

如果细胞膜对于多种离子都具有通透性，如 K^+、Na^+ 和 Cl^- 离子都可以通过，那么，在稳态平衡状态下，各种离子的 Nernst 电位可以分别用式（11.12）～式（11.14）计算。不过，膜电位 $V_\mathrm{m} = v_i - v_o$ 是由所有离子共同形成的，并且受到各种离子浓度和通透性的影响。本节先介绍两种离子通透的计算方法，下一节再介绍任意数目离子通透的情况。

假设细胞膜对 K^+ 和 Cl^- 两种离子具有通透，但是对体积较大的阳离子 R^+ 不通透，如图 11.8 所示，平衡状态下，K^+ 和 Cl^- 这两种离子的 Nernst 电位必须相等，即 $E_\mathrm{K} = E_\mathrm{Cl}$，也就是

$$E_\mathrm{K} = \frac{KT}{q}\ln\frac{[\mathrm{K}^+]_o}{[\mathrm{K}^+]_i} = E_\mathrm{Cl} = \frac{KT}{q}\ln\frac{[\mathrm{Cl}^-]_i}{[\mathrm{Cl}^-]_o} \tag{11.15}$$

简化后，可得

$$\frac{[\mathrm{K}^+]_o}{[\mathrm{K}^+]_i} = \frac{[\mathrm{Cl}^-]_i}{[\mathrm{Cl}^-]_o} \tag{11.16}$$

此方程就是著名的 Donnan 平衡方程。由此可得空间电荷平衡原则，也就是在给定空

间中阳离子数目与阴离子数目相等，呈电中性。这样，在平衡状态下，离子的跨膜扩散仍然存在，只是每个跨膜流动的 K^+ 必须伴随着一个 Cl^-，从而满足空间电荷呈中性的要求。如果图 11.8 中不存在 R^+，那么，平衡时，膜内外两侧的 K^+ 和 Cl^- 离子浓度将相等。如果细胞外液中存在 R^+，则细胞膜两侧的 KCl 浓度就不相等，下面举例说明这种情况。

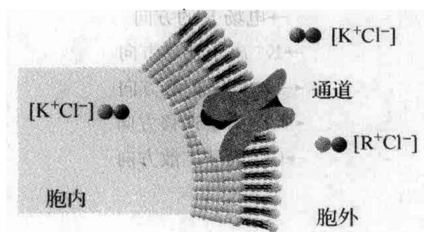
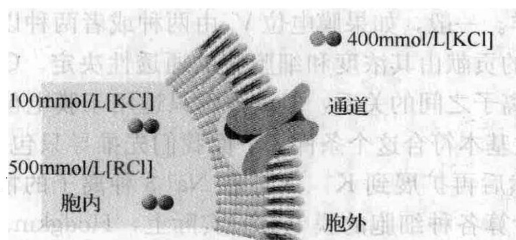


图 11.8 细胞膜对 K^+ 和 Cl^- 都具有通透，但是对较大的阳离子 R^+ 不通透

例 11.1 假设细胞膜对 K^+ 和 Cl^- 通透，但对较大的阳离子 R^+ 不通透。请计算以下初始条件下稳态时的平衡离子浓度。



解：由质量守恒定律可得

$$[K^+]_i + [K^+]_o = 500$$
$$[Cl^-]_i + [Cl^-]_o = 1000$$

再由空间电荷平衡原则可得

$$[K^+]_i + 500 = [Cl^-]_i$$
$$[K^+]_o = [Cl^-]_o$$

根据 Donnan 平衡方程，有

$$\frac{[K^+]_o}{[K^+]_i} = \frac{[Cl^-]_i}{[Cl^-]_o}$$

将质量守恒方程中求得的$[K^+]_o$和$[Cl^-]_o$代入 Donnan 平衡方程可得

$$\frac{500-[K^+]_i}{[K^+]_i} = \frac{[Cl^-]_i}{1000-[Cl^-]_i}$$

然后将电荷平衡方程中求得的 $[Cl^-]_i$代入上式，得到

$$\frac{500-[K^+]_i}{[K^+]_i} = \frac{[K^+]_i+500}{1000-[K^+]_i-500} = \frac{[K^+]_i+500}{500-[K^+]_i}$$

求解此方程可以得到，稳态时$[K^+]_i = 167\text{mmol/L}$。再由质量守恒方程和电荷平衡方程求得稳态时$[K^+]_o = 333\text{mmol/L}$，$[Cl^-]_i = 667\text{mmol/L}$，$[Cl^-]_o = 333\text{mmol/L}$。在稳态和室温条件下，两种离子的 Nernst 电位都是 18mV，以$[K^+]$ 为例，即

$$E_K = v_i - v_o = 26 \ln\frac{333}{167} = 18\text{mV}$$

下图显示了稳态时的情况：

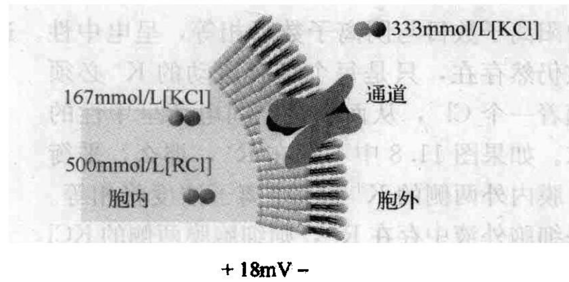
→电场 E 的方向

→K^+离子迁移方向

←K^+离子扩散方向

←Cl^-离子迁移方向

→Cl^-离子扩散方向



■

11.4.4 Goldman 方程

枪乌贼巨神经纤维轴突的膜静息电位是－60mV，这个电位与 Na^+ 或 K^+ 的 Nernst 电位都不一样。一般，如果膜电位 V_m 由两种或者两种以上离子共同形成，那么，每种离子对 V_m 的贡献由其浓度和细胞膜的通透性决定。Goldman 方程定量描述了 V_m 与各种可通透离子之间的关系，但该方程只适用于膜电位（即电场）恒定的情况，细胞膜静息电位基本符合这个条件。下面我们先推导只包含 K^+ 和 Cl^- 两种离子的 Goldman 方程，然后再扩展到 K^+、Cl^- 和 Na^+ 3 种离子的情况。Goldman 方程被生理学家广泛用于计算各种细胞的膜电位，实际上，Hodgkin、Huxley 和 Katz 当年在研究枪乌贼巨神经纤维轴突时就用到了这个方程。

假设细胞膜如图 11.9 所示。为了计算 K^+ 和 Cl^- 两种离子所形成的膜电位 V_m，首先独立推导每种离子在恒定电场下的净流量方程，然后再应用空间电荷平衡原则，就可以求得 Goldman 方程。

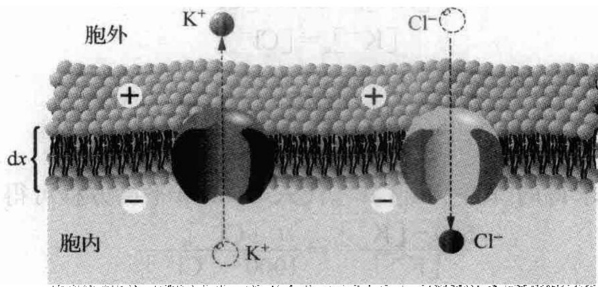


图 11.9　对 K^+ 和 Cl^- 两种离子通透的细胞膜示意图（膜的厚度 $dx=\delta$）

钾离子

设 K^+ 的迁移率为 μ_K，则净流量方程为

$$J_K=-\frac{KT}{q}\mu_K\frac{d[K^+]}{dx}-\mu_K Z_K[K^+]\frac{dv}{dx} \tag{11.17}$$

在恒定电场下，有

$$\frac{dv}{dx}=\frac{\Delta v}{\Delta x}=\frac{V}{\delta} \tag{11.18}$$

将此式代入式（11.17），且 $Z_K=1$，可得

$$J_K=-\frac{KT}{q}\mu_K\frac{d[K^+]}{dx}-\mu_K Z_K[K^+]\frac{V}{\delta} \tag{11.19}$$

令 K^+ 的通透率为 P_K，则

$$P_K=\frac{\mu_K KT}{\delta q}=\frac{D_K}{\delta} \tag{11.20}$$

于是，将上式代入式（11.19）可得

$$J_K=-\frac{P_K q}{KT}V[K^+]-P_K\delta\frac{d[K^+]}{dx} \tag{11.21}$$

重组该方程可得

$$dx=\frac{d[K^+]}{\frac{-J_K}{P_K\delta}-\frac{qV[K^+]}{KT\delta}} \tag{11.22}$$

此式两边同时求积分，并假设 J_K 与 x 无关，可得

$$\int_0^\delta dx=\int_{[K^+]_i}^{[K^+]_o}\frac{d[K^+]}{\frac{-J_K}{P_K\delta}-\frac{qV[K^+]}{KT\delta}} \tag{11.23}$$

结果是

$$x\Big|_0^\delta=\frac{KT\delta}{qV}\ln\left(\frac{J_K}{P_K\delta}+\frac{qV[K^+]}{KT\delta}\right)\Bigg|_{[K^+]_i}^{[K^+]_o} \tag{11.24}$$

即

$$\delta=-\frac{KT\delta}{qV}\ln\left\{\frac{\frac{J_K}{P_K\delta}+\frac{qV[K^+]_o}{KT\delta}}{\frac{J_K}{P_K\delta}+\frac{qV[K^+]_i}{KT\delta}}\right\} \tag{11.25}$$

将该方程两边的 δ 消去，并将 $-\frac{KT}{qV}$ 移到方程的另一边，然后两边同时取指数可得

$$e^{-\frac{qV}{KT}}=\frac{\frac{J_K}{P_K\delta}+\frac{qV[K^+]_o}{KT\delta}}{\frac{J_K}{P_K\delta}+\frac{qV[K^+]_i}{KT\delta}} \tag{11.26}$$

求解此方程可得 J_K 为

$$J_K=\frac{qVP_K}{KT}\left(\frac{[K^+]_o-[K^+]_i e^{\frac{-qV}{KT}}}{e^{\frac{-qV}{KT}}-1}\right) \tag{11.27}$$

氯离子

Cl^- 的推导过程与 K^+ 一样，最终可得

$$J_{Cl}=\frac{qVP_{Cl}}{KT}\left(\frac{[Cl^-]_o e^{\frac{-qV}{KT}}-[Cl^-]_i}{e^{\frac{-qV}{KT}}-1}\right) \tag{11.28}$$

式中　P_{Cl}——Cl^- 的通透率。

钾离子和氯离子的总和效应

由电荷平衡原则可知 $J_K = J_{Cl}$，联合方程（11.27）和（11.28）可得

$$P_K\left([K^+]_o - [K^+]_i e^{\frac{-qV}{KT}}\right) = P_{Cl}\left([Cl^-]_o e^{\frac{-qV}{KT}} - [Cl^-]_i\right) \tag{11.29}$$

求解其中的指数项可得

$$e^{\frac{-qV}{KT}} = \frac{P_K[K^+]_o + P_{Cl}[Cl^-]_i}{P_K[K^+]_i + P_{Cl}[Cl^-]_o} \tag{11.30}$$

再求解 V 可得

$$V = v_o - v_i = -\frac{KT}{q}\ln\left(\frac{P_K[K^+]_o + P_{Cl}[Cl^-]_i}{P_K[K^+]_i + P_{Cl}[Cl^-]_o}\right) \tag{11.31}$$

写为膜电位 V_m，即

$$V_m = \frac{KT}{q}\ln\left(\frac{P_K[K^+]_o + P_{Cl}[Cl^-]_i}{P_K[K^+]_i + P_{Cl}[Cl^-]_o}\right) \tag{11.32}$$

该方程就是 Goldman 方程。由于 Na^+ 对膜电位也很重要，因此，同理可以推导出包含 K^+、Cl^- 和 Na^+ 3 种离子的 Goldman 方程为

$$V_m = \frac{KT}{q}\ln\left(\frac{P_K[K^+]_o + P_{Na}[Na^+]_o + P_{Cl}[Cl^-]_i}{P_K[K^+]_i + P_{Na}[Na^+]_i + P_{Cl}[Cl^-]_o}\right) \tag{11.33}$$

式中 P_{Na}——Na^+ 的通透率。

为了推导方程（11.33），首先求 J_{Na}，然后利用电荷平衡原则，即 $J_K + J_{Na} = J_{Cl}$，就可以得到此方程。一般，如果某种离子的通透率比其他离子都大得多，那么，由 Goldman 方程计算得到的膜电位 V_m 就很接近这种离子的 Nernst 电位。

表 11.1 和表 11.2 分别列出了枪乌贼巨神经纤维和蛙骨骼肌细胞内外的几种主要离子浓度、它们的相对通透率和 Nernst 电位。由于枪乌贼巨神经纤维比较粗大，并且没有髓鞘，实验操作比较容易，因此实验研究中用得很多。通常，脊椎动物神经元胞内和胞外的离子浓度是枪乌贼巨神经纤维的 1/4～1/3。

表 11.1 枪乌贼巨神经纤维细胞内外几种重要离子的大概浓度、相对通透率和 Nernst 电位

离子	胞内浓度/(mmol/L)	胞外浓度/（mmol/L)	相对通透率	Nernst 电位/mV
K^+	400	20	1	−74
Na^+	50	440	0.04	55
Cl^-	52	560	0.45	−60

注：此处的通透率是相对值，即 P_K：P_{Na}：P_{Cl}，而不是绝对值。表中数据都在 6.3℃下测得，KT/q 约等于 25.3mV。

表 11.2 蛙骨骼肌细胞内外几种重要离子的大概浓度、相对通透率和 Nernst 电位

离子	胞内浓度/(mmol/L)	胞外浓度/（mmol/L)	相对通透率	Nernst 电位/mV
K^+	140	2.5	1.0	−105
Na^+	13	110	0.019	56
Cl^-	3	90	0.381	−89

注：表中数据在室温下测得，KT/q 约等于 26mV。

例 11.2 请计算枪乌贼巨神经纤维在 6.3℃时的跨膜电位 V_m。

解：利用方程（11.33）以及表 11.1 所提供的数据，可得

$$V_m = 25.3 \times \ln\left(\frac{1\times20+0.04\times440+0.45\times52}{1\times400+0.04\times50+0.45\times560}\right)\text{mV} = -60\text{mV} \quad ■$$

11.4.5 离子泵

静息时，细胞膜两侧各种电荷和离子浓度必须保持恒定不变，也就是说，流进细胞的电荷量必须与流出细胞的电荷量相等，否则跨膜电位 V_m 就会产生变化。此时，在浓度梯度和电场梯度的共同作用下，Na^+ 离子会流向胞内；而 K^+ 离子的扩散作用力大于迁移作用力，因而 K^+ 离子会流向胞外。空间电荷平衡原则要求 Na^+ 的内流量等于 K^+ 的外流量。虽然这些离子流量能够相互抵消从而维持电荷平衡，但是，不可能只有这种单方向的离子流动。否则，$[K^+]_i$ 最终会趋于 0，而 $[Na^+]_i$ 则会一直增加，根据 Goldman 方程，也就会导致 V_m 的变化。

细胞膜上的 Na-K 泵可以防止 K^+ 和 Na^+ 离子浓度梯度发生变化，这种离子泵以恒定的速度将 Na^+ 泵出细胞并将 K^+ 泵入细胞。将 Na^+ 泵出细胞是逆浓度梯度和电场梯度进行的，是一种需要消耗能量的主动运输。图 11.10 是 Na-K 泵以及主动离子通道和被动离子通道的示意图。

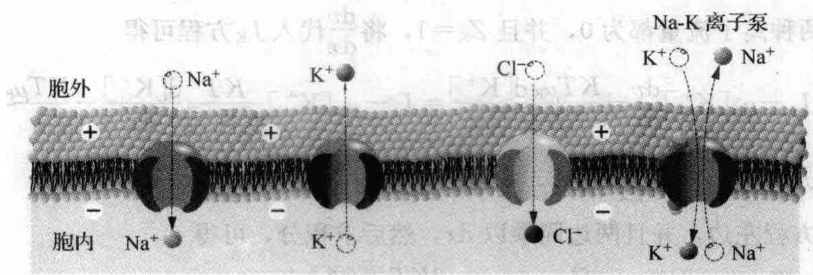


图 11.10 Na-K 主动离子泵以及主动离子通道和被动离子通道示意图

Na-K 泵其实会产生电，也就是它会跨膜传输净电荷。不产生电的离子泵运行时不导致净电荷的传输。但是，很多神经元的 Na-K 离子泵每次从胞外移入 2 个 K^+ 时就会从胞内移出 3 个 Na^+，这使得 V_m 比只存在被动通道时计算所得的值要稍微负一些。

一般，当细胞膜处于静息状态时，主动离子流量和被动离子流量相平衡，并且膜两侧可以维持稳定的电位，但必须满足以下 2 个条件：

1）细胞膜对某些离子不通透；

2）细胞膜上存在主动离子泵。

Na-K 主动离子泵的存在使得 K^+ 和 Na^+ 离子的浓度可以保持不变，从而迫使 V_m 处于某个特定的电位值。其他离子的浓度则由 V_m 决定。例如，由于 Cl^- 离子只能通

过被动通道穿越细胞膜，因此，静息时细胞内外 Cl^- 浓度之比就可以用 Nernst 方程计算，此时 $E_{Cl}=V_m$，即

$$\frac{[Cl^-]_i}{[Cl^-]_o}=e^{\frac{qV_m}{KT}} \tag{11.34}$$

例 11.3　假设细胞膜上存在主动 K^+ 离子泵以及 K^+ 和 Cl^- 的被动离子通道，并且初始时膜两侧的 KCl 浓度不相等。请写出主动 K^+ 离子泵流量的表达式。

解：根据空间电荷平衡原则，$J_{Cl}=J_K$，而

$$J_K=J_p-\frac{KT}{q}\mu_K\frac{d[K^+]}{dx}-\mu_K Z_K[K^+]\frac{dv}{dx}$$

$$J_{Cl}=-\frac{KT}{q}\mu_{Cl}\frac{d[Cl^-]}{dx}-\mu_{Cl}Z_{Cl}[Cl^-]\frac{dv}{dx}$$

式中　J_p——主动 K^+ 离子泵的流量。

由 J_{Cl}方程求解$\frac{dv}{dx}$，且 $Z_{Cl}=-1$，可得

$$\frac{dv}{dx}=\frac{KT}{q[Cl^-]}\frac{d[Cl^-]}{dx}$$

根据空间电荷平衡原则有 $[Cl^-]=[K^+]$，代入以上方程可得

$$\frac{dv}{dx}=\frac{KT}{q[K^+]}\frac{d[K^+]}{dx}$$

平衡时，两种离子流量都为 0，并且 $Z_K=1$，将$\frac{dv}{dx}$代入 J_K方程可得

$$J_K=0=J_p-\mu_K[K^+]\frac{dv}{dx}-\frac{KT\mu_K}{q}\frac{d[K^+]}{dx}=J_p-\mu_K[K^+]\frac{KT}{q[K^+]}\frac{d[K^+]}{dx}-\frac{KT\mu_K}{q}\frac{d[K^+]}{dx}$$

$$=J_p-\frac{2KT\mu_K}{q}\frac{d[K^+]}{dx}$$

将 J_p移到方程左边，并且两边同乘以 dx，然后求积分，可得

$$-\int_0^\delta J_p dx=-\frac{2KT\mu_K}{q}\int_{[K^+]_i}^{[K^+]_o}d[K^+]$$

即

$$J_p=\frac{2KT\mu_K}{q\delta}([K^+]_o-[K^+])_i$$ ■

注意：本例中如果没有离子泵，那么，平衡时细胞膜两侧的离子浓度将相等。

11.5　细胞膜的等效电路模型

本节中，我们将用前面所介绍的方法建立细胞膜的等效电路模型，它是下一节将要讲述的 Hodgkin-Huxley（HH）动作电位模型的基础，HH 模型另外引入了电压依赖性和时间依赖性离子通道。正如 11.3 节和 11.4 节所述，神经元具有 3 种被动电特性：电动势、电阻和电容。神经细胞膜是一层脂双分子层，其上穿插着各种各样

的离子通道。这些离子通道要么是一直开放的被动通道；要么是门控的主动通道，可以开放和关闭。每种离子通道都具有选择通透性。另外，细胞膜上还有维持跨膜电位 V_m 的 Na-K 主动离子泵。

11.5.1 细胞膜的电动势、电阻和电容特性

电动势特性

如图 11.5 所示，静息时，3 种主要离子 K^+、Na^+ 和 Cl^- 在细胞膜两侧的分布不均匀，它们通过被动离子通道跨膜流动。膜两侧电荷分离，并形成了可用 Goldman 方程（11.33）计算的膜电位 V_m。

在每种离子特异性通道两侧都存在着离子的浓度梯度，从而产生了电动势，驱使离子以恒定的速度穿越通道。该离子的 Nernst 电位就是通道两侧的电位差，很容易用电池来模拟这个电位差，图 11.11 显示了 K^+ 离子通道的模型。Na^+ 和 Cl^- 离子通道也可以用类似的模型来模拟，只是电池的电动势要用它们自己的 Nernst 电位。

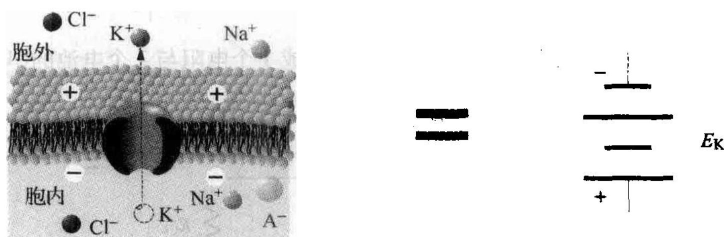


图 11.11　用电池模拟 K^+ 离子通道的电动势，其值等于 K^+ 离子的 Nernst 电位

注：根据常规定义，胞外作为地电位。由表 11.1 可知，K^+ 的 Nernst 电位为负，与图中所示电池的极性相反，驱使 K^+ 流出细胞。

电阻特性

除了电动势之外，细胞膜上的每个离子通道都存在电阻，它会阻碍电荷穿越通道，这主要是由于离子通过通道时与通道壁产生碰撞，使能量转化为热能。通道电阻一般用电导 G 来表示，单位为西门子（S），其含义是离子通过时的容易程度。由于各个通道的电导是并联的，因此总的细胞膜电导就是离子通道总数 N 乘以每个离子通道的电导 G'，即

$$G=N\times G'$$

有时把电导写成电阻 $R=\frac{1}{G}$ 更方便，其单位是欧姆（Ω）。

如图 11.12 所示，现在可以用一个电阻与一个电池的串联作为一种离子通道的等效电路。

细胞膜电导的大小与其通透性有关，但是在生理学意义上，两者并不能相互替代。电导与细胞膜的状态有关，并且随着离子浓度的变化而变化，还与跨膜离子流量成正比。通透性则描述了细胞膜对于某种特定离子的状态。例如，如果细胞膜两

侧没有任何离子存在，那么，在电压作用下就没有离子流过细胞膜，无论细胞膜上有多少离子通道开放，其电导都是 $G=0$。而此时离子通透性却保持不变。

3 种离子的等效电路

如图 11.12 所示，K^+、Na^+ 和 Cl^- 中每一种离子通道都可以用各自的 Nernst 电位和电阻串联的等效电路来表示，将这 3 个等效电路合成 1 个电路，然后将细胞外液和细胞内液都用短路连线表示，就形成了细胞膜静息状态下的电路模型（见图 11.13）。

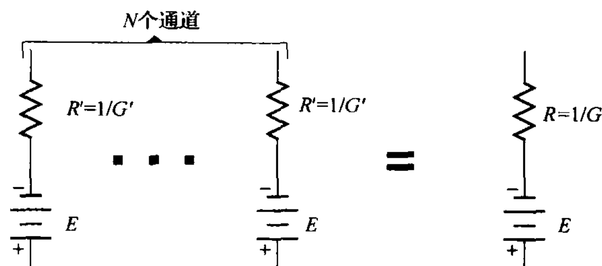


图 11.12　N 个离子通道的等效电路可以组合成 1 个电阻与 1 个电池的串联电路

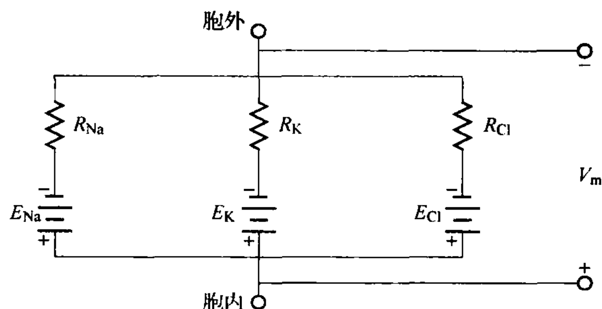
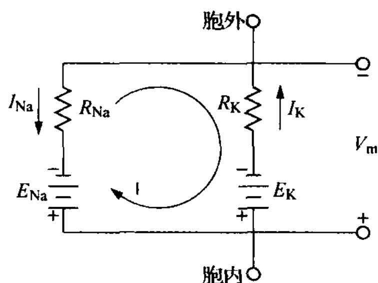


图 11.13　静息时小块神经纤维细胞膜上的被动离子通道模型，其中每种离子通道等效为 1 个电阻与 1 个电池的串联

例 11.4　请计算蛙骨骼肌细胞膜（参见表 11.2）的膜电位 V_m，假设 Cl^- 通道忽略不计，并且 $R_K=1.7\ k\Omega$，$R_{Na}=15.67\ k\Omega$。



解：细胞膜等效电路如下，假设回路电流为 I，I_{Na} 为通过钠离子通道的电流，I_K 为通过钾离子通道的电流。根据本书第 8 章介绍的电路分析理论可得：

$$E_{Na}+IR_{Na}+IR_K-E_K=0$$

求解 I，即

$$I=\frac{E_K-E_{Na}}{R_{Na}+R_K}=\frac{(-105-56)\times10^{-3}}{(15.67+1.7)\times10^3}=-9.27\mu A$$

于是

$$V_m=E_{Na}+IR_{Na}=-89\text{mV}$$

注意，应该有 $I=-I_{Na}$，$I=-I_K$，即 $I_{Na}=I_K$。生理学上，这表示内向的 Na^+ 电流恰好与外向的 K^+ 电流相等。■

例 11.5 请计算蛙骨骼肌细胞膜的膜电位 V_m，假设 $R_{Cl}=3.125\ \text{k}\Omega$。

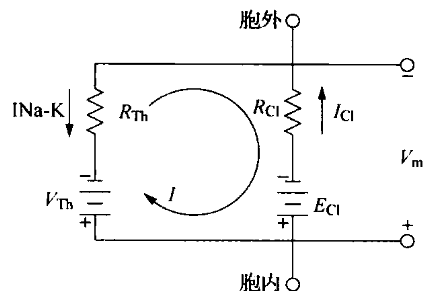
解： 为了求解本题，首先求出例 11.4 所示电路的戴维南等效电路。由于

$$\text{等效电压}\quad V_{Th}=V_m=-89\text{mV}$$

并且

$$\text{等效电阻}\quad R_{Th}=\frac{R_{Na}\times R_K}{R_{Na}+R_K}=1.534\ \text{k}\Omega$$

于是戴维南等效电路如右图所示。



由表 11.2 可知 $E_{Cl}=V_{Th}$，因此没有电流流过 Cl^- 离子通道。这是一般神经元的真实情况，其膜电压由 K^+ 离子和 Na^+ 离子的相对电导以及它们的 Nernst 电位决定。Nernst 电位则由维持离子浓度梯度的自动离子泵确定。Cl^- 离子通常只是被动地分布于膜两侧。■

Na-K 泵

在上述例 11.4 和前面 11.4 节中已经介绍过，即使在静息状态下细胞膜也会有稳定的外向 K^+ 电流和内向 Na^+ 电流。如果没有任何其他措施，任其一直流下去，那么，最终就会使得 E_K 和 E_{Na} 都趋于 0。为了防止这种情况发生，"电流源" Na-K 泵产生的电流被用来抵消这些被动电流，Na-K 泵的电路模型如图 11.14 所示。

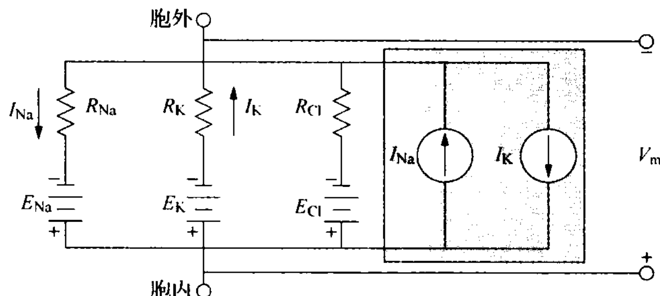


图 11.14 静息时小块神经纤维上的 3 种被动离子通道的电路模型，各种离子通道等效为 1 个电阻和 1 个电池的串联电路，Na-K 主动离子泵则用阴影框内的两个电流源来模拟

11.5.2 电容特性

任何由绝缘材料隔开的两个导电体之间都会产生电容。神经元的细胞内液和细胞外液就是导电体，而脂双分子层是绝缘材料（见图 11.2），神经元细胞膜的电容约为 $1\mu F/cm^2$。膜电容的存在表明，除非通过离子通道传输，否则离子不可能通过细胞膜。

将细胞膜的膜电容与电动势、电阻结合起来，细胞膜的电路模型就变成图 11.15 所示。加入膜电容之后，膜电压的变化就不再是瞬间的变化，而是具有一阶时间常数的指数变化。为了考察电容的效应，用戴维南等效电路将图 11.15 的电路简化为图 11.16 所示的电路，其中等效电阻 R_{Th} 和电压 V_{Th} 为

$$R_{Th}=\frac{1}{\frac{1}{R_K}+\frac{1}{R_{Na}}+\frac{1}{R_{Cl}}} \tag{11.35}$$

$$V_{Th}=-\frac{R_{Na}R_{Cl}E_K+R_KR_{Cl}E_{Na}+R_KR_{Na}E_{Cl}}{R_{Na}R_{Cl}+R_KR_{Cl}+R_KR_{Na}} \tag{11.36}$$

该细胞膜电路模型的时间常数就是 $\tau=R_{Th}\times C_m$，在 5τ 时间之后响应可以达到稳态值的 99%以上。一般神经元的 τ 值范围为 1～20ms。另外，稳态时，电容相当于开路，于是 $V_{Th}=V_m$。

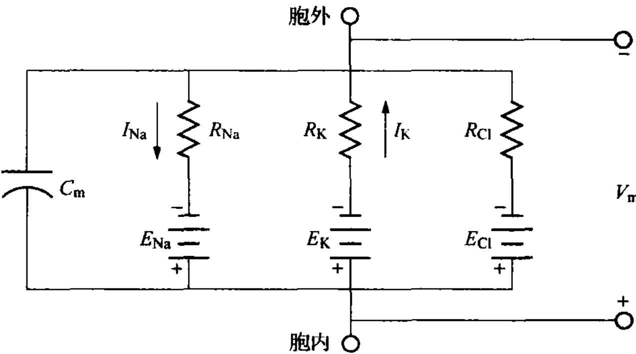


图 11.15 静息时小块神经纤维细胞膜的电路模型

注：模型中包含了所有被动电特性。图 11.14 中的 Na-K 主动离子泵被去除，因为它对电路没有作用。

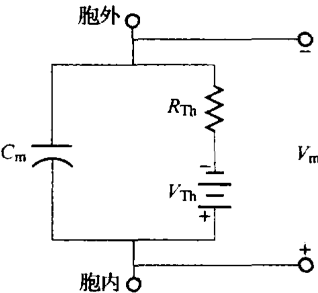
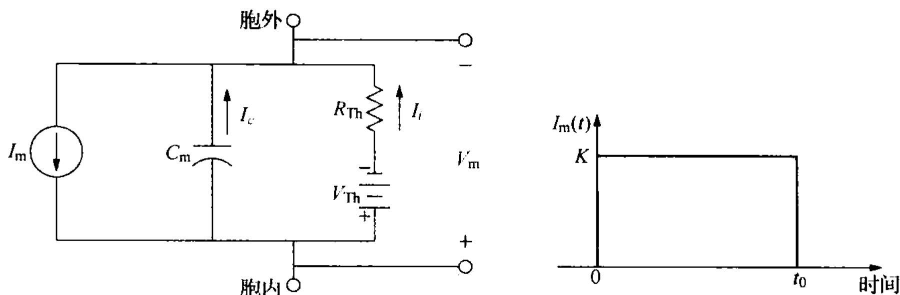


图 11.16 图 11.15 所示电路模型的戴维南等效电路

例 11.6　请计算由跨膜电流脉冲引起的膜电压 V_m 的变化。



解：实际实验操作时，如上图所示刺激电流脉冲通过胞内电极和胞外电极施加在细胞膜上。设电流脉冲 I_m 的幅值为 K，持续时间为 t_0，并在 $t=0$ 时刻发生，那么，由此引起的膜电压 V_m 可以用基尔霍夫电流定律计算，即

$$-I_m+\frac{V_m-V_{Th}}{R_{Th}}+C_m\frac{dV_m}{dt}=0$$

该节点电流方程的拉普拉斯变换为

$$-I_m(s)+\frac{V_m(s)}{R_{Th}}-\frac{V_{Th}}{sR_{Th}}+sC_mV_m(s)-C_mV_m(0^+)=0$$

合并同类项之后可得

$$\left(s+\frac{1}{C_mR_{Th}}\right)V_m(s)=V_m(0^+)+\frac{I_m(s)}{C_m}+\frac{V_{Th}}{sC_mR_{Th}}$$

电流脉冲的拉普拉斯变化为 $I_m(s)=\frac{K}{s}(1-e^{-t_0s})$，将此 $I_m(s)$ 代入以上节点方程并重组各项，可得

$$V_m(s)=\frac{V_m(0^+)}{\left(s+\frac{1}{C_mR_{Th}}\right)}+\frac{K(1-e^{-t_0s})}{sC_m\left(s+\frac{1}{C_mR_{Th}}\right)}+\frac{V_{Th}}{sC_mR_{Th}\left(s+\frac{1}{C_mR_{Th}}\right)}$$

将上式进行部分分式展开，并注意 $V_m(0^+)=V_{Th}$，可得

$$V_m(s)=KR_{Th}\left[\frac{1}{s}-\frac{1}{\left(s+\frac{1}{C_mR_{Th}}\right)}\right](1-e^{-t_0s})+\frac{V_{Th}}{s}$$

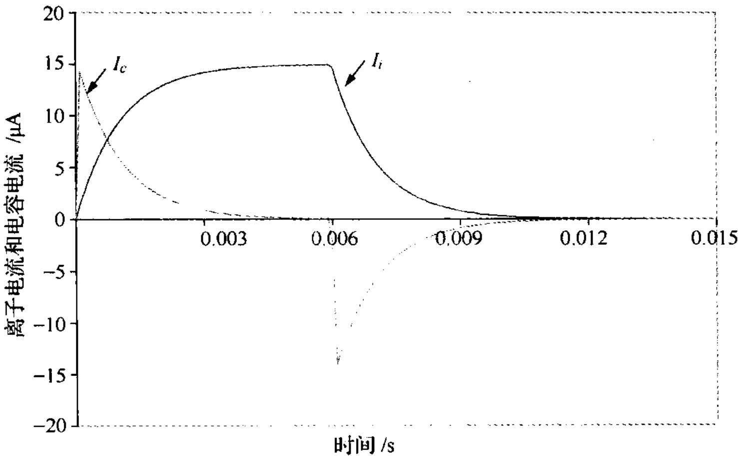
将上式变换到时域，可得

$$V_m(t)=V_{Th}+R_{Th}K\left(1-e^{-\frac{t}{R_{Th}C_m}}\right)u(t)-R_{Th}K\left(1-e^{-\frac{t-t_0}{R_{Th}C_m}}\right)u(t-t_0)$$

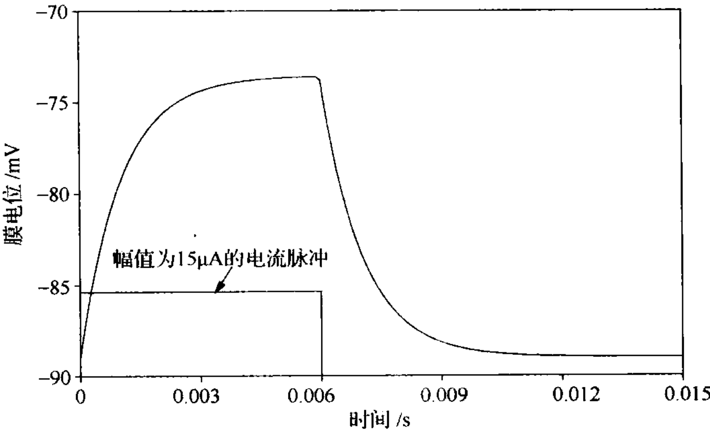
离子电流 I_i 和电容电流 I_c 如下图所示，即

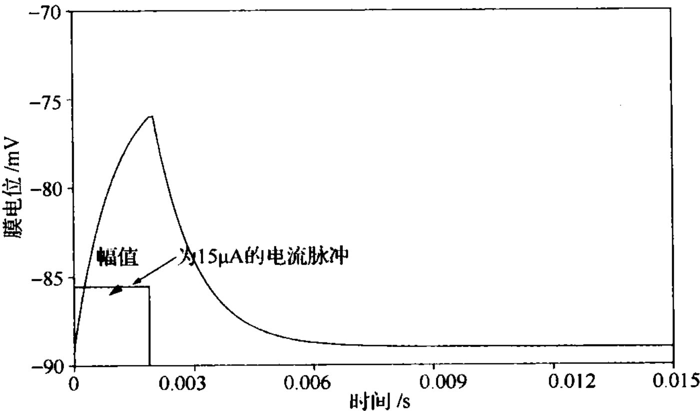
$$I_i=\frac{V_m-V_{Th}}{R_{Th}}$$

$$I_c=C_m\frac{dV_m}{dt}$$



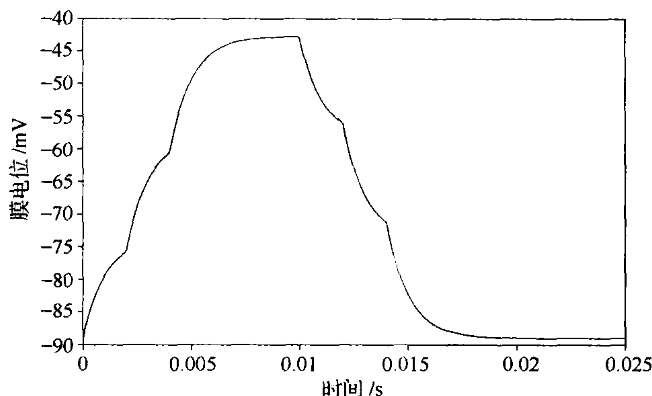
下面两图所示是持续时间分别为 6s（上图）和 2s（下图）的 15μA 电流脉冲引起的膜电压 V_m 的响应，所用参数为蛙骨骼肌细胞参数，时间常数约为 1ms。注意，上图的膜电压 V_m 在电流脉冲回到 0 之前就已经达到稳态；而下图的 V_m 未达到稳态值。





在突触部位，时间常数的大小对于突触电流的整合非常重要。每一个神经递质小泡的释放都相当于一个电流脉冲。注意，时间常数越大，膜兴奋的持续时间就越长。一般各个不同刺激引起的兴奋都不会同步发生，但是由于存在时间常数 τ，相当数量的刺激响应可以叠加起来，从而产生较大的膜电压变化，最终可以诱发产生动作电位传输信号。

下图所示是一系列持续时间为 6s 的 15μA 电流脉冲所产生的 V_m 响应的叠加，这些电流脉冲发生的时刻分别为 0、2、4、6、8ms。由于后一个脉冲发生在前一个脉冲的 5 τ 时间之内，因此各个脉冲的 V_m 响应叠加在一起，从而使膜的去极化接近 −45mV。如果脉冲之间的时间间隔大于 5 τ，则 V_m 将是一系列分开的独立脉冲响应。



■

11.5.3 膜电压随距离的变化

图 11.15 和图 11.16 所示的电路模型只描述了一小块或者一小段细胞膜。例 11.6 显示了注入细胞膜的电流脉冲会引起膜电压 V_m 的变化，但是，小块膜上局部 V_m 的变化会产生流向膜的邻近区域的电流，这些电流又会引起邻近区域膜电压 V_m 的变化，依此类推，V_m 的变化就会扩散到整个细胞膜。由于神经细胞树突内部的空间比胞外空间要小得多，因此，树突内流动的电流所受到的阻抗要比胞外空间电流受到的阻抗大得多。树突直径越大，电流在树突内传播时的电阻就越小。如图 11.17 所示，为了模拟这种效应，树突各个分段之间加上了体现胞内阻抗的电阻 R_a。该模型其实是在 x、y、z 3 个方向上都连续的三维细胞膜表面模型。胞外空间较大，因此胞外阻抗忽略不计，只用短路连线来表示。

如图 11.18 所示，如果在树突的某个分段注入电流，就像例 11.6 那样，只是这里的电流脉冲持续时间 t_0 没有例 11.6 那么大。那么，稳态时，电容 C_m 引起的暂态响应已经过去，细胞膜电阻和树突内电阻就决定了电流的分布。显然，电流注入处所在的那个分段流出细胞膜的电流最大，因为与其他分段相比，电流流经的电阻最小，

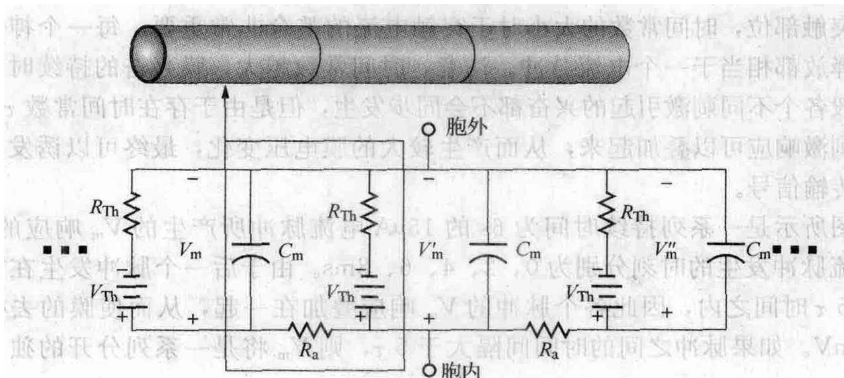


图 11.17 通过轴向电阻 R_a 连接起来的一系列树突细胞膜分段的等效电路

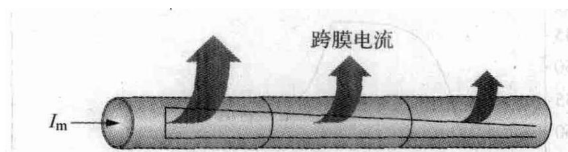


图 11.18 稳态时树突细胞膜的跨膜电流示意图

注：由于电流总是沿最小电阻路径流动，大部分电流在刺激电流脉冲注入处流出细胞膜，随着与该注入处的距离增大，跨膜流出的电流逐渐减小。

只有 R_{Th}；从紧邻的下一个分段流出细胞膜的电流稍小一些，此电流需流经电阻 $R_{Th}+R_a$。因此，电流注入处之外各个分段的膜电压变化 ΔV_m 与 C_m 无关，只取决于 R_{Th} 和 R_a 的相对值。距离电流注入处第 n 个分段流出细胞膜的电流需经过的电阻为 $R_{Th}+n\times R_a$。可见，距离电流注入处越远的分段，流出细胞膜的电流越小，膜电压变化 ΔV_m 也就越小，因为 ΔV_m 等于流出该分段细胞膜的电流乘以 R_{Th}。膜电压的变化 ΔV_m 随着距离的增加以指数形式衰减，其计算公式为

$$\Delta V_m = V_o e^{\frac{-x}{\lambda}} \tag{11.37}$$

式中 $\lambda=\sqrt{\frac{R_{Th}}{R_a}}$——细胞膜空间常数；

x——膜电压计算位置与电流脉冲注入处之间的距离；

V_o——电流脉冲注入处膜电压的变化量。

λ 的取值范围为 0.1～1mm，λ 越大，电流脉冲刺激沿树突轴向能够传播的距离就越远。

11.6 动作电位的 Hodgkin-Huxley 模型

1952 年，Hodgkin 和 Huxley 接连发表了 5 篇论文，报道了枪乌贼巨轴突神经纤维中进行的一系列实验，以及由此推导出来的一种动作电位模型。前 4 篇论文阐述

了各种实验研究，说明了动作电位期间细胞膜发生的变化，最后一篇论文则建立了数学模型。这个模型并不是根据本章所介绍的定律和理论推导出来的生理模型，而是用指数函数拟合实验数据曲线得到的经验模型。下面就介绍 Hodgkin 和 Huxley 所做实验的主要部分以及这个经验模型，这里所用的曲线图都是根据 HH 模型用 SIMULINK 程序仿真得到，采用的参数是枪乌贼巨轴突的数据。

11.6.1 动作电位和电压钳实验

神经元具有传递动作电位的能力，这使得信号可以在神经系统中长距离传播，动作电位的一个重要特性就是在远距离传导过程中其幅值不会衰减。当轴丘（见图 11.1）处的膜电压 V_m 达到某个阈值时就会诱发神经元产生动作电位。一旦 V_m 达到阈值，Na^+ 和 K^+ 主动门控离子通道的电导就会随时间和膜电压发生变化，驱使 V_m 趋向于 Na^+ 离子的 Nernst 电位 E_{Na}，紧接着又趋向于 K^+ 离子的 Nernst 电位 E_K，最后再回到静息电位。Hodgkin 和 Huxley 最先报道了这种离子通道的电导变化规律，Katz 也参与了部分工作，并且是其中一篇论文的作者之一。图 11.19 所示是一个典型的动作电位波形，其阈值电位约为 -40mV。

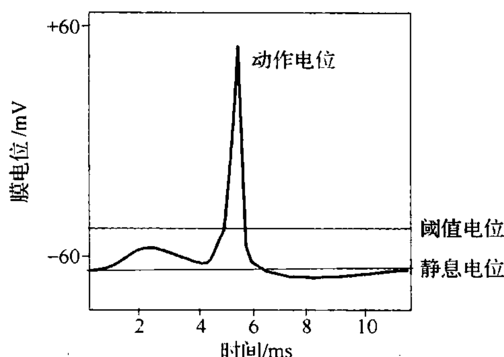


图 11.19 动作电位示意图（膜电压大约在 5ms 时达到阈值电位）

注：动作电位的起因是离子通道的电导产生了电压和时间依赖性的变化，其上升支由 Na^+ 离子通道电导的变化引起，而下降支则由 K^+ 离子通道电导变化引起。

要使轴丘的膜电压 V_m 达到阈值电位，树突和胞体部位必须接收足够多的突触后膜输入刺激。如前所述，距离轴丘越远，突触后膜上接受的刺激对轴丘处 V_m 变化的贡献就越小。并且，由于细胞膜具有时间常数，突触后膜的刺激与由其导致的轴丘 V_m 变化之间总是存在延时。因此，在描述轴丘 V_m 的分级应答时，时间和距离是两个重要的参数。

一旦 V_m 达到阈值，主动 Na^+ 离子通道随即开放，导致 Na^+ 的内流，使细胞膜进一步去极化。这种去极化又会使 Na^+ 电导增大，结果产生更多的 Na^+ 内流。如图 11.20 所示，这种再生过程迅速使 V_m 趋近 E_{Na}，直到 Na^+ 通道失活为止。K^+ 离子通道也发生类似的电导变化过程，并驱使 V_m 重新回到静息电位，但是 K^+ 电导的变化

较为缓慢。一旦动作电位被启动之后，它就会自行完成整个过程，这被称为“全或无”现象。Na^+和K^+主动门控通道的电导变化都是膜电压V_m和时间的函数。

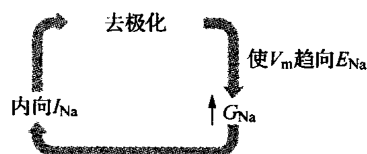


图 11.20 Na^+通道电导变化示意图

动作电位沿轴突高速传播，并且在有髓鞘的神经纤维中是在郎飞氏节之间跳跃式传播。这是因为髓鞘的膜电容很小，使细胞膜接近纯电阻，从而使V_m的变化几乎可以瞬时完成。

为了研究动作电位，Hodgkin 和 Huxley 选用了无髓鞘的枪乌贼巨轴突作为研究材料，这种神经纤维的直径很大，约为 1mm，并且在 6.3℃海水中可以存活数小时。当时已有的学说认为，动作电位是由于细胞膜通透性的巨大变化引起的，这种通透性的变化使所有离子都可以自由流过细胞膜，从而驱使V_m趋于 0。然而，Hodgkin 和 Huxley 的实验结果并不是这样。他们与哥伦比亚大学的 Cole 和 Curtis 合作，在研究中使用了两种新的实验技术：空间钳位和电压钳位，这是他们取得成功的关键。下面介绍他们的实验技术。

Hodgkin 和 Huxley 将银丝插入轴突，消除了胞内电阻R_a，从而使一长段轴突细胞膜上的V_m都能够保持同样的值，这就是空间钳位。而电压钳位技术则用于控制膜电压V_m，使得膜通透性改变时I_{Na}内流和I_K外流不能引起进一步的膜去极化。选择枪乌贼巨神经纤维作为实验材料具有两个优点：其直径大且可以在海水中长时间存活；它只有两种电压和时间依赖性的离子通道。其他神经元的电压—时间依赖性的离子通道都超过两种，很难进行分析，甚至不可能分析。

电压钳

为了研究K^+和Na^+电压—时间依赖性通道的电导变化过程，Hodgkin 和 Huxley 用电压钳将电压和时间这两个机制分开，在各个恒定不变的膜电压下只研究通道的时间依赖性变化过程。图 11.21 所示是利用前述等效电路表示的电压钳实验，其中K^+和Na^+通道用电压—时间依赖性的可变电阻表示，而Na^+、K^+和Cl^-的被动通道用漏电阻R_l表示（即被动通道的戴维南等效电路）。如图 11.22 所示，电压钳的作用是中断Na^+和K^+通道电阻与膜电位之间的偶联。如前所述，如果膜电位不钳位，那么Na^+和K^+通道电阻的变化会引起膜电位的变化，然后又进一步改变Na^+和K^+通道的电阻，如此反复。

如图 11.23 所示，电压钳用两对电极实现。实验时，一对电极注入电流I_m，用以维持膜电压V_m恒定不变；另一对电极则用来检测V_m。为了计算Na^+通道和K^+通道的电导，实验时还需测量I_m。图 11.21 中显示了测量V_m和I_m的电表，这些电表放在海水浴之外。如今，测量仪器的输出信号都用模/数转换器进行采样，然后将采样数据存入计算机硬盘用于分析处理。但是，在 1952 年，这些测量仪器的输出只能用纸带记录仪记录。施加钳位电压V_c之后，Na^+通道电导发生变化，产生Na^+离子内流，欲驱使膜电位进一步上升，超过V_c；但此时电压钳会通过膜电流I_m的变化去

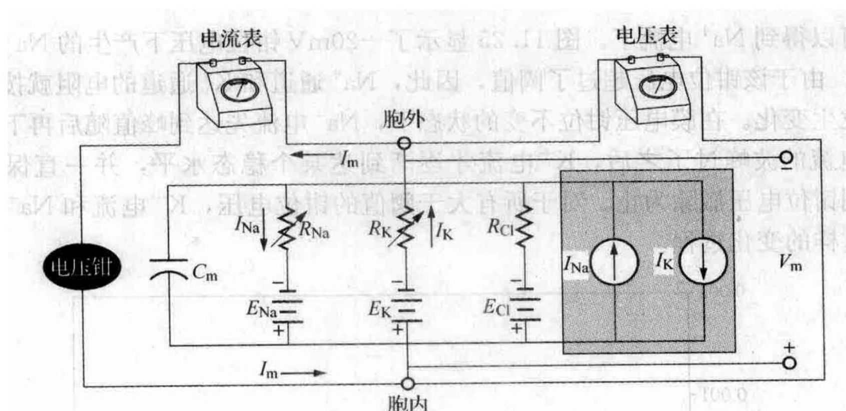


图 11.21　电压钳位下的枪乌贼巨轴突无髓鞘纤维的等效电路模型

注：K^+ 和 Na^+ 通道用电压—时间依赖性的可变电阻表示，Na^+、K^+ 和 Cl^- 被动离子通道用漏电阻 R_l 表示。Na-K 泵用阴影部分的电路来表示。实验时，细胞膜浸浴于海水之中。

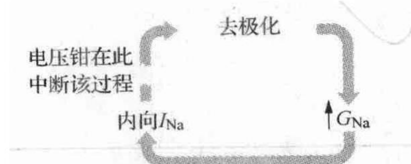


图 11.22　电压钳实验中断图 11.20 所示的再生循环过程

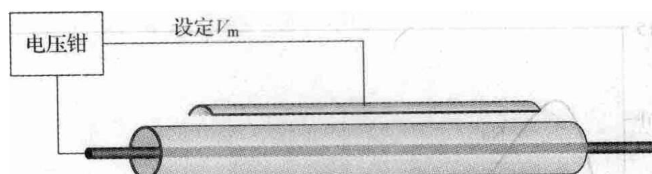


图 11.23　电压钳实验的物理装置

除细胞内侧增加的正电荷，从而保持膜电压 V_m 不变。电压钳实验中，电流 I_m 是因变量，钳位电压 V_c 则是自变量。

实验时，先设定某个钳位电压 V_c，然后记录将膜电压 V_m 维持在该钳位电压时所经历的膜电流 I_m 的变化过程。图 11.24 显示了由 −20mV 钳位电压引起的 I_m。开始时，V_m 的阶跃变化引起一个较大的跨膜电流，该电流主要是膜电容电流。钳位电压同时产生一个恒定的跨膜漏电流，其值等于

$$I_l=\frac{V_c-E_l}{R_l} \tag{11.38}$$

将此电容电流和漏电流减去之后，剩下的就是 Na^+ 电流和 K^+ 电流。为了进一步将 Na^+、K^+ 两个电流分开，Hodgkin 和 Huxley 在胞外溶液中用一种不可通透的较大的阳离子取代 Na^+ 离子，这样就去除了 Na^+ 电流，只剩下 K^+ 电流了。然后，恢复胞外溶液中的 Na^+ 离子，并将记录得到的总的 I_m 减去电容电流、漏电流和 K^+ 离子电

流，就可以得到 Na^+ 电流了。图 11.25 显示了－20mV 钳位电压下产生的 Na^+ 电流和 K^+ 电流。由于该钳位电压超过了阈值，因此，Na^+ 通道和 K^+ 通道的电阻就按照一定的模式发生变化。在膜电压钳位不变的状态下，Na^+ 电流先达到峰值随后再下降。等到 Na^+ 电流的波峰过了之后，K^+ 电流才逐渐到达某个稳态水平，并一直保持这个值，直到钳位电压撤除为止。对于所有大于阈值的钳位电压，K^+ 电流和 Na^+ 电流都要经历这样的变化过程。

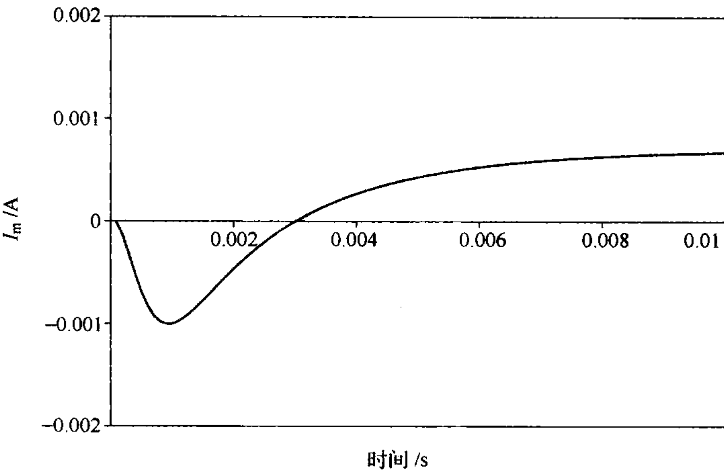


图 11.24 由－20mV 钳位电压引起的膜电流 I_m 的变化

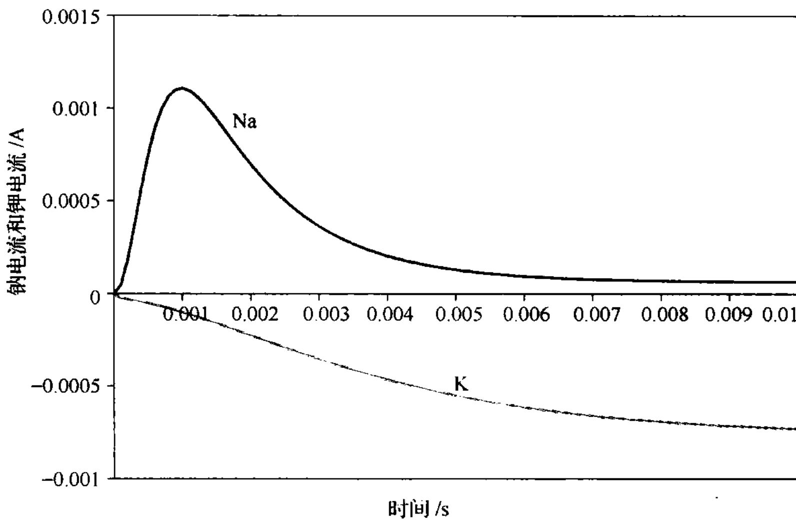


图 11.25 由－20mV 钳位电压产生的 Na^+ 电流和 K^+ 电流

在图 11.21 的电路中应用欧姆定理，并利用图 11.25 的电流数据，可以方便地求取 Na^+ 通道和 K^+ 通道的电导（即电阻）：

$$I_K = \frac{V_m - E_K}{R_K} = G_K(V_m - E_K) \tag{11.39}$$

$$I_{Na} = \frac{E_{Na} - V_m}{R_{Na}} = G_{Na}(E_{Na} - V_m) \tag{11.40}$$

式中 G_K、G_{Na}——K^+、Na^+通道的电导。

图 11.26 所示就是在−50～20mV 的各个钳位电压下，利用以上两个公式计算得到的两种离子通道电导随时间变化的曲线。

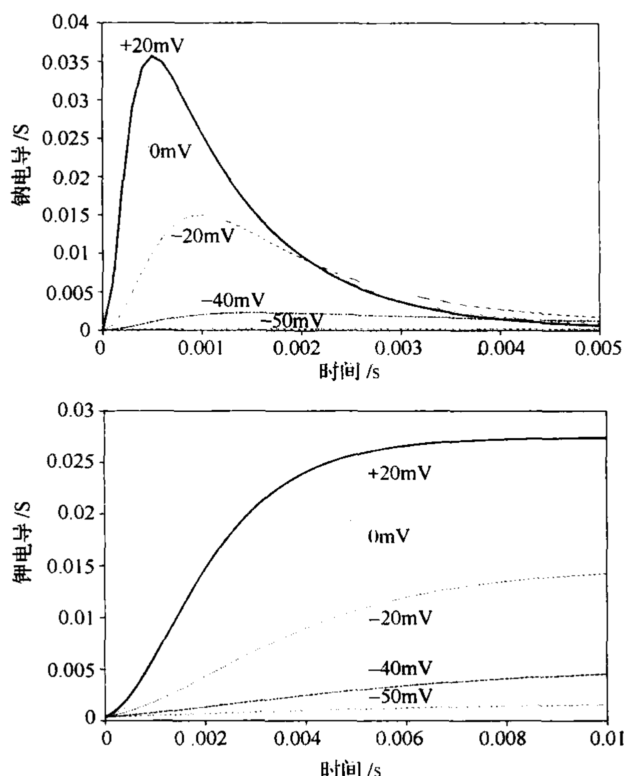


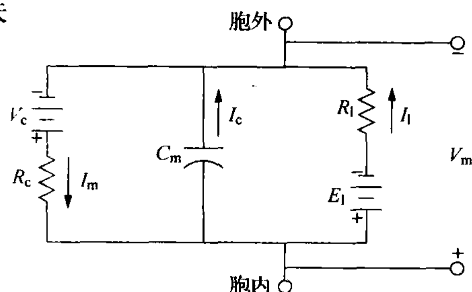
图 11.26 在−50mV（阈值下）到+20mV 的各个钳位电压作用下 Na^+ 电导和 K^+ 电导的变化曲线

注：两张图的时间坐标不一样。

对于所有阈上钳位电压，Na^+ 通道开放的速度都比 K^+ 通道快，并且一段时间之后 Na^+ 通道随即失活；而 K^+ 通道则不同，只要钳位电压保持不变，K^+ 通道就一直保持开放状态。一旦 Na^+ 通道失活，除非细胞膜已经复极化回到静息电位，否则 Na^+ 通道不能再开放。Na^+ 通道处于失活状态的这段时间被称为不应期（refractory period）。如果电压钳位在 Na^+ 通道的整个变化过程结束之前就被撤除，那么，Na^+ 通道的电导 G_{Na} 几乎立即回到 0，而 K^+ 通道的电导 G_K，无论 Na^+ 通道的变化过程是否完成，都会逐渐回到 0。

例 11.7 请计算阈下钳位电压作用下的跨膜电容电流 I_c 和漏电流 I_l。

解： 由于刺激是阈下的，不妨设 Na^+ 和 K^+ 的电压一时间依赖性离子通道没有被激活，因此，可以忽略这些通道（实际上并非如此，详见例 11.9）。这样，细胞膜的等效电路可以表示为右图所示。



其中，R_s 是连接线的电阻。应用基尔霍夫电流定律可得

$$C_m\frac{dV_m}{dt}+\frac{V_m-E_l}{R_l}+\frac{V_m-V_c}{R_s}=0$$

重组该方程可得

$$C_m\frac{dV_m}{dt}+\frac{R_l+R_s}{R_lR_s}V_m=\frac{R_lV_c+R_sE_l}{R_lR_s}$$

由初始条件 $V_m(0)=E_l$，可得如下解

$$V_m=\frac{R_lV_c+R_sE_l}{R_l+R_s}+\frac{R_l(E_l-V_c)}{R_l+R_s}e^{\frac{-(R_l+R_s)t}{R_lR_sC_m}}$$

从而

$$I_c=C_m\frac{dV_m}{dt}=\frac{E_l-V_c}{R_s}e^{\frac{(R_l+R_s)t}{R_lR_sC_m}}$$

$$I_l=\frac{V_m-E_l}{R_l}$$

稳态时

$$I_l=\frac{V_c-E_l}{R_l+R_s}$$

■

动作电位的重建

通过分析不同幅值和不同持续时间电压钳位脉冲作用下计算得到的 G_{Na} 和 G_K，Hodgkin 和 Huxley 建立了一组描述动作电位的非线性方程，这些方程的仿真结果可以精确地描述各种不同刺激诱发的动作电位响应。在学习这些方程之前，有必要根据上述实验数据，先定性了解动作电位发生期间各种离子通道的变化过程。动作电位由阈值上去极化触发，去极化使 G_{Na} 增加并产生 Na^+ 离子内流；Na^+ 电流进一步使细胞膜去极化，并使 Na^+ 内流增大。这个过程驱使 V_m 不断接近 Na^+ 的 Nernst 电位 E_{Na}。如图 11.26 所示，G_{Na} 是时间和膜电压的函数，它达到最大值后又迅速回落。在 G_{Na} 回落期间，G_K 一直在增加，并使细胞膜超极化，驱使 V_m 从 E_{Na} 变到 E_K。G_K 的增加形成了 K^+ 外流，而 K^+ 外流电流进一步使膜超极化，这又增加了 K^+ 外流。这个过程驱使 V_m 不断趋于 K^+ 的 Nernst 电位 E_K，E_K 值低于静息电位。图 11.27 显示了动作电位期间 V_m、G_{Na} 和 G_K 的变化过程。

图 11.16 的等效电路很有用，可以仿真较小的阈下去极化状态下细胞膜的情况。该电路模型假设 K^+ 电流和 Na^+ 电流都很小，可以忽略不计。由例 11.6 可知，在这种情况下，注入细胞的电流脉冲上升沿快速形成一个呈指数衰减的电容电流，以及一个呈指数增加的漏电流 I_l。电流脉冲结束时（即下降沿），再次产生电容电流并以

指数形式衰减到 0，漏电流也以指数形式衰减为 0。

随着外加的刺激电流脉冲幅值的增加，细胞膜的去极化也增加，从而激活了 Na^+、K^+ 电压一时间依赖性通道。当去极化足够大时，内流的 Na^+ 电流会超过外流的 K^+ 电流与漏电流之和（即 $I_{Na}>I_K+I_l$），此时的膜电压 V_m 就是阈值。一旦细胞膜的去极化达到了阈值，Na^+ 和 K^+ 的电压一时间依赖性通道就被启动，随即完成图 11.27 所示的整个变化过程。

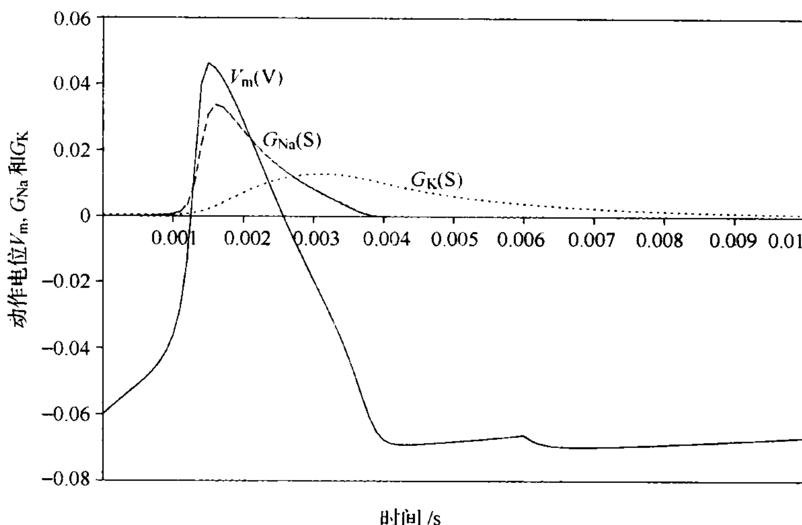


图 11.27　动作电位期间 V_m、G_{Na}和 G_K 的变化过程

如果用缓慢上升的刺激电流使细胞膜去极化，则阈值会变得比较高。因为在缓慢接近阈值的过程中，未达到阈值之前就会发生部分 G_{Na}通道的失活，而 G_K 通道则开始激活，因此，要满足 $I_{Na}>I_K+I_l$ 的条件，V_m 值要比快速接近阈值时大很多。

11.6.2　钠离子和钾离子的电导方程

Hodgkin 和 Huxley 用如下形式的数学方程建立 G_{Na}和 G_K模型：

$$G(t)=(A+B\mathrm{e}^{-Ct})^{D} \tag{11.41}$$

参数 A、B、C 和 D 的值用枪乌贼巨轴突的电压钳实验数据来估计。从该方程似乎看不出通道电导的电压依赖性，其实有些参数是随电压变化的。两种离子通道的电导方程中 D 都设定为 4，这样，方程与实验数据拟合最好。图 11.27 实际上是用 SIMULINK 仿真程序以及 Hodgkin 和 Huxley 给出的参数计算得到的，SIMULINK 是 MATLAB 的一个仿真软件包。本节后面会介绍该仿真程序的具体内容。下面先分析 G_{Na}和 G_K的数学方程。

钾离子通道的电导方程

钾电导的变化曲线如图 11.26 所示，在外加电脉冲刺激期间，电导曲线逐渐上升并趋于最大值，这个过程用以下 G_K方程来描述：

$$G_K = \overline{G}_K n^4 \tag{11.42}$$

式中 $\overline{G}_K$——最大 K^+ 电导；

n——可以看作速率常数。

n 是如下微分方程的解：

$$\frac{dn}{dt} = \alpha_n(1-n) - \beta_n n \tag{11.43}$$

其中

$$\alpha_n = 0.01\frac{V+10}{e^{\frac{V+10}{10}}-1}$$

$$\beta_n = 0.125e^{\frac{V}{80}}$$

$$V = V_{rp} - V_m$$

式中 V_{rp}——没有任何刺激情况下细胞膜的静息电位。

注意，V 是膜电压偏离静息电位的值，应该是负值。显然，由方程（11.43）可知，G_K是时间依赖性变量，同时也是电压依赖性变量，因为 n 中含有电压依赖性参数 α_n 和 β_n。

钠离子通道的电导方程

钠电导变化的曲线如图 11.26 所示，曲线先上升到最大值然后下降。这个变化过程可以用 G_{Na}的模型方程来描述，该模型是两个函数的乘积，其中一个描述曲线的上升相，另一个则描述下降相，该模型为

$$G_{Na} = \overline{G}_{Na} m^3 h \tag{11.44}$$

式中 $\overline{G}_{Na}$——最大 Na^+ 电导；

m、h——可以看作速率常数。

m 和 h 是以下微分方程的解：

$$\frac{dm}{dt} = \alpha_m(1-m) - \beta_m m \tag{11.45}$$

$$\frac{dh}{dt} = \alpha_h(1-h) - \beta_h h \tag{11.46}$$

其中

$$\alpha_m = 0.1\frac{V+25}{e^{\frac{V+25}{10}}-1}$$

$$\beta_m = 4e^{\frac{V}{18}}$$

$$\alpha_h = 0.07e^{\frac{V}{20}}$$

$$\beta_h = \frac{1}{e^{\frac{V+30}{10}}+1}$$

注意，m 描述了 G_{Na} 的上升相，而 h 描述了下降相。式（11.43）、式（11.45）和式（11.46）中 α_i 和 β_i 的单位都是 m/s；n、m 和 h 都无量纲，且取值范围为 0～1。

例 11.8 请用 Hodgkin-Huxley 模型计算枪乌贼巨轴突在静息电位处的 G_K 和 G_{Na}，参数为 $\overline{G}_K = 36\times10^{-3}$S，$\overline{G}_{Na} = 120\times10^{-3}$S。

解： 静息时，G_K和G_{Na}都是常数，其值由n、m和h决定。由于细胞膜处于稳态，因此，$\frac{dn}{dt}=0$，$\frac{dm}{dt}=0$，$\frac{dh}{dt}=0$。应用式（11.43）、式（11.45）和式（11.46），在静息和稳态情况下，有

$$n=\frac{\alpha_n^0}{\alpha_n^0+\beta_n^0}$$

$$m=\frac{\alpha_m^0}{\alpha_m^0+\beta_m^0}$$

$$h=\frac{\alpha_h^0}{\alpha_h^0+\beta_h^0}$$

其中，α_i^0是$V=0$时的$\alpha(i=n, m, h)$；β_i^0是$V=0$时的$\beta(i=n, m, h)$。经计算可得$\alpha_n^0=0.0582$，$\beta_n^0=0.125$，$n=0.31769$，$\alpha_m^0=0.2236$，$\beta_m^0=4$，$m=0.05294$，$\alpha_h^0=0.07$，$\beta_h^0=0.04742$，$h=0.59615$。因此，在静息和稳态时

$$G_K=\overline{G}_K n^4=36.0\times10^{-3}(0.31769)^4=0.3667\times10^{-3}\text{S}$$

$$G_{Na}=\overline{G}_{Na}m^3h=120.0\times10^{-3}(0.05294)^3\times0.59615=0.010614\times10^{-3}\text{S}$$ ■

11.6.3 膜电压的时间依赖性方程

图 11.28 所示是外部刺激I_m作用下的细胞膜模型，适用于动作电位的仿真。应用基尔霍夫定律可得

$$I_m=G_K(V_m-E_K)+G_{Na}(V_m-E_{Na})+\frac{(V_m-E_l)}{R_l}+C_m\frac{dV_m}{dt} \quad (11.47)$$

其中G_K和G_{Na}是由式（11.42）和式（11.44）给出的电压—时间依赖性电导。

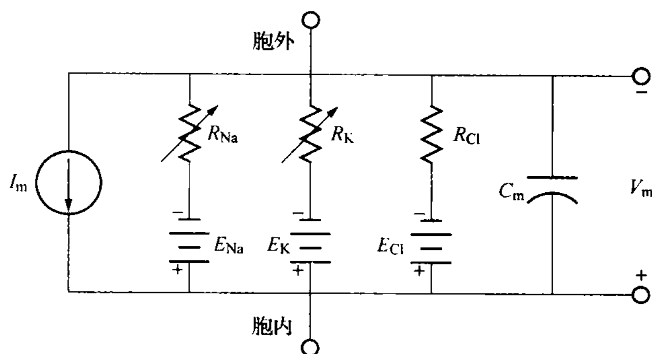


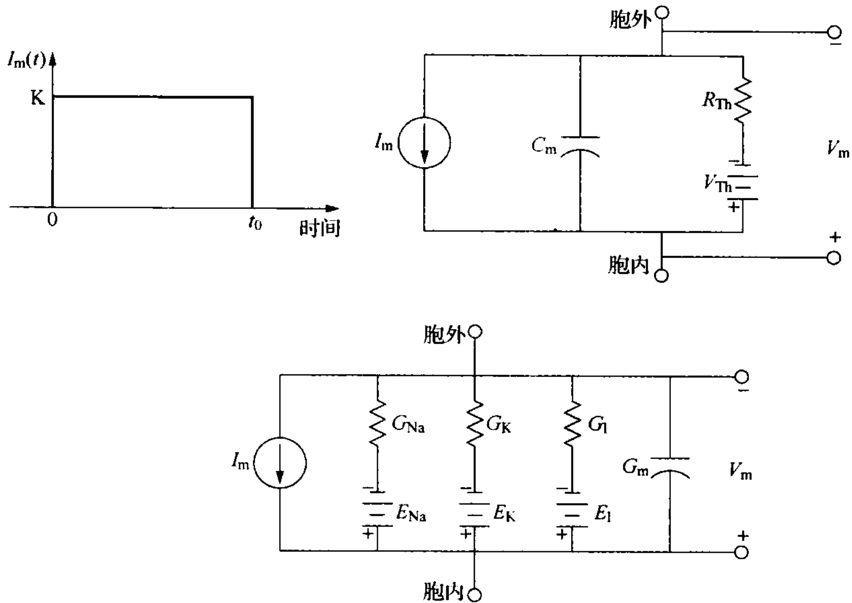
图 11.28 一段无髓鞘枪乌贼巨轴突的电路模型

注：K^+通道和Na^+通道的电压—时间依赖性可变电导（即电阻）由式（11.42）和式（11.44）给定。Na^+、K^+和Cl^-的被动离子通道用漏电阻R_l和 Nernst 电位E_l表示。Na-K 泵没有画出，它不对电路其他部分提供电流。

例 11.9 如果要使枪乌贼巨轴突的膜电位从-60mV 上升到-40mV、然后再恢复到静息电位，请计算所需外加电流脉冲的幅值和宽度。K^+、Na^+电导的变化忽略

不计，用静息时的 G_K 和 G_{Na} 值进行计算。并且，枪乌贼巨轴突的 Hodgkin-Huxley 的参数为 $G_l=\frac{1}{R_l}=0.3\times10^{-3}\text{S}$，$\overline{G}_K=36\times10^{-3}\text{S}$，$\overline{G}_{Na}=120\times10^{-3}\text{S}$，$E_K=-72\times10^{-3}\text{V}$，$E_l=-49.4\times10^{-3}\text{V}$，$E_{Na}=55\times10^{-3}\text{V}$，$C_m=1\times10^{-6}\text{F}$。

解： 假设电流 I_m 如下图所示：



在前面的例 11.8 中已求得 K^+、Na^+ 离子通道静息时的电导分别为 $G_K=0.3667\times10^{-3}\text{S}$，$G_{Na}=0.010614\times10^{-3}\text{S}$。在本题的阈下刺激电流作用下，$G_K$ 和 G_{Na} 保持不变，因此，图 11.28 所示电路可以简化为如上所示的电路。为了方便计算，用戴维南等效电路简化该电路，等效电路的电阻为

$$R_{Th}=\frac{1}{G_{Na}+G_K+G_l}=1.4764\ \text{k}\Omega$$

且等效电压 $V_{Th}=-60\text{mV}$。

于是，本题的解与例 11.6 的解一样，即

$$V_m(t)=V_{Th}+R_{Th}K(1-e^{-\frac{t}{R_{Th}C_m}})u(t)-R_{Th}K(1-e^{-\frac{t-t_0}{R_{Th}C_m}})u(t-t_0)$$

为了方便计算，假设电流脉冲宽度 $t_0>5\tau$，因此，对于 $t\leqslant t_0$，$V_m=-40\text{mV}$ 是本题的稳态值，于是上式成为

$$V_m=-0.040=V_{Th}+K\times R_{Th}=-0.060+K\times1476.4$$

由此可得电流脉冲幅值 $K=13.6\mu\text{A}$。由于 $\tau=R_{Th}C_m=1.47\text{ms}$，当电流脉冲幅值为 $K=13.6\mu\text{A}$ 时，任何宽度 t_0 大于 $5\tau=7.35\text{ms}$ 的电流脉冲都可以将膜电压 V_m 升至 -40mV。当然，如果要诱发动作电位，就需要更大幅值的电流脉冲，正如前面 11.6.1 节末尾所述，因为随着 V_m 以指数形式接近阈值（只是无限接近而不能达到），Na^+ 通道会被激活且随即失活。 ■

为了计算动作电位期间的膜电压 V_m，需要求解 4 个微分方程（式（11.43）、式（11.45）、式（11.46）和式（11.47））以及 6 个代数方程（式（11.43）、式（11.45）和式（11.46）中的 α_i 和 β_i）。由于这些方程组成的方程组中存在 n^4 和 m^3 的电导项，是非线性方程组，因此求其解析解是不可能的。要求解 V_m，就需要用数值方法。用于求解非线性系统数值解的计算机工具有很多，本书将使用 MATLAB 的通用工具箱 SIMULINK，它可以方便地用于求解线性和非线性、连续和离散动态系统的数值解，是一个广泛流行的仿真软件，具有友好的用户界面，完全集成于 MATLAB 之中。SIMULINK 是交互式的，其程序的开发基于系统模块框图，在一般计算机平台上都可以运行。

动作电位的 SIMULINK 仿真程序如图 11.29～11.32 所示，其设计过程介绍如下。首先，求方程（11.47）中的最高阶导数，可得下式：

$$\frac{dV_m}{dt}=\frac{1}{C_m}(I_m+G_K(E_K-V_m)+G_{Na}(E_{Na}-V_m)+G_l(E_l-V_m)) \tag{11.48}$$

根据此式，就可以创建系统模块图，也就是用积分器、求和器等建立 SIMULINK 程序。

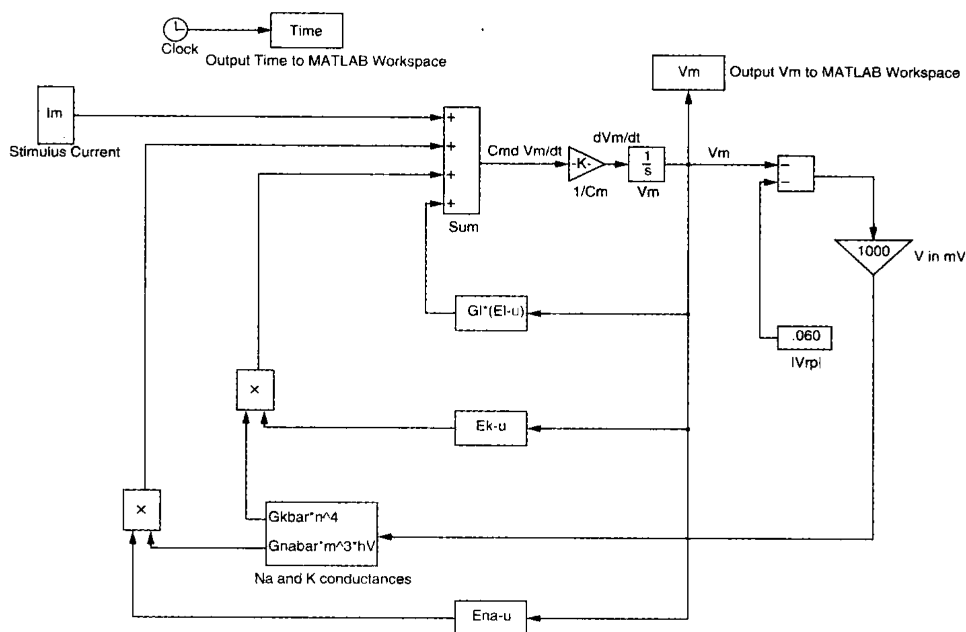


图 11.29　SIMULINK 动作电位仿真程序的主模块图

注：刺激电流脉冲由图 11.30 所示的两个阶跃函数相减得到。Na^+、K^+ 电导的函数模块如图 11.31 和 11.32 所示。

图 11.29 所示是主模块图，图 11.30～11.32 是子系统模块图。其中的“工作空间”输出模块用于将仿真结果传送到 MATLAB 环境下作曲线图。仿真所用参数值取

自 Hodgkin 和 Huxley 的实验结果，即 $G_l=\frac{1}{R_l}=0.3\times10^{-3}$S，$\overline{G}_K=36\times10^{-3}$S，$\overline{G}_{Na}=120\times10^{-3}$S，$E_K=-72\times10^{-3}$V，$E_l=-49.4\times10^{-3}$V，$E_{Na}=55\times10^{-3}$V，$C_m=1\times10^{-6}$F。图 11.27 就是该 SIMULINK 程序的动作电位仿真结果。模块 Gl* (El-u)、Ek-u 和 Ena-u 分别表示式（11.48）中的 $G_l(E_l-V_m)$、(E_K-V_m) 和 $(E_{Na}-V_m)$ 这几项。

刺激电流脉冲用图 11.30 所示的 SIMULINK 两个阶跃函数构成。其中一个阶跃函数始于 $t=0$ 时刻，幅值为 K；另一个始于 $t=t_0$ 时刻，幅值为 $-K$。电流脉冲要能够快速将 V_m 升至阈值以上水平，才能诱发动作电位。

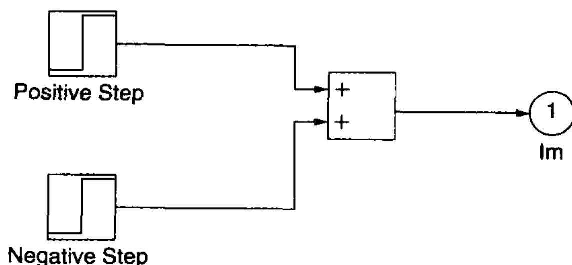


图 11.30 刺激电流脉冲的 SIMULINK 程序

图 11.31 所示是表示 Na^+ 通道和 K^+ 通道电导的 SIMULINK 程序。函数模块 Gkbar * u^4、Gnabar * u^3 和 Gnabar * m^3 * h 分别表示 $\overline{G}_K n^4$、$\overline{G}_{Na}m^3$ 和 $\overline{G}_{Na}m^3h$。子系统 n、m 和 h 如图 11.32 所示，它们分别基于 6 个代数方程（即式（11.43）、式（11.45）和式（11.46）中的 α_i 和 β_i）。

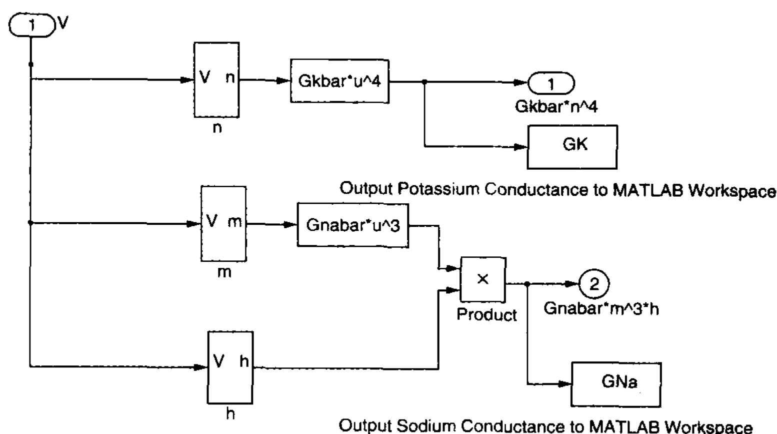


图 11.31 K^+、Na^+ 通道电导的 SIMULINK 程序

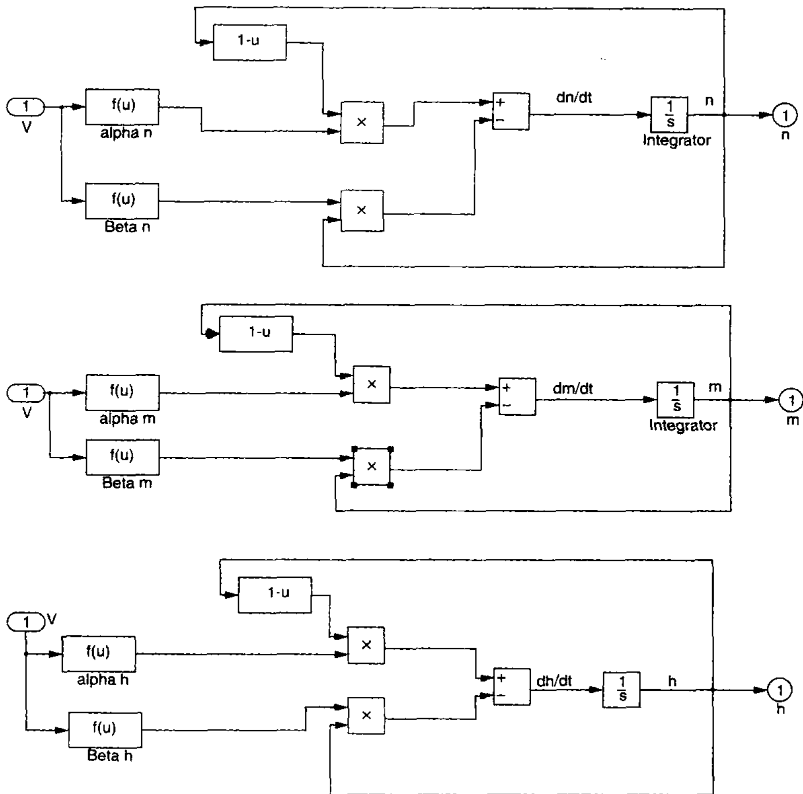


图 11.32 式（11.43)、式（11.45）和式（11.46）的α_i和β_i项的 SIMULINK 程序

11.7 神经元整体模型

本节将介绍神经元的整体模型，包括树突、胞体、轴突以及突触前末梢。正如本章 11.5.3 节所述，树突和轴突可以用一组串联的圆柱形房室模型来模拟，各个房室之间用轴向电阻连接。轴突和树突都与胞体相连。当然，真实的神经元有很多不同的结构，例如，树突连接到轴突，轴突又与胞体相连等。但神经元的基本组成是一个胞体、一条轴突加多个树突。注意，树突和轴突不一定是直径恒定的圆柱体，它们在向外延伸的过程中会变得越来越细。

图 11.17 所示是树突的一般房室模型，只包含被动通道；图 11.28 所示则是轴突的房室模型，其中包含了轴丘和郎飞氏节部位的主动通道。轴突的髓鞘部分可以用类似树突的房室模型来模拟，其中只包含电容、被动离子通道和轴向电阻等被动特性。图 11.33 所示就是一段有髓鞘轴突的模型，其中有 3 个被动房室，以及一个表示郎飞氏节的主动房室，此结构可以扩展成任意数目房室的组合。此外，根据神经元的类型，胞体可以用主动房室或者被动房室来模拟。

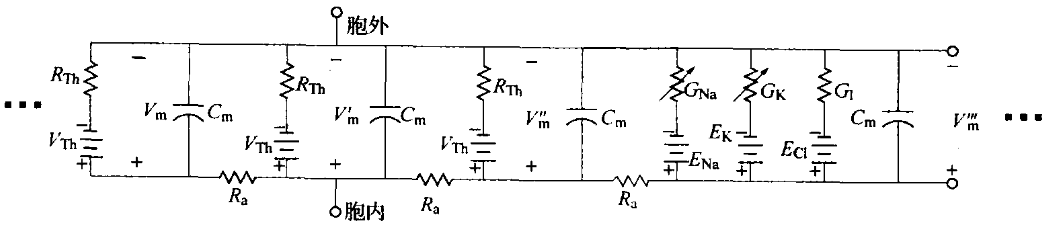


图 11.33 包含主动房室和被动房室的一段轴突模型

应用基尔霍夫电流定律，图 11.33 所示的轴突模型的表达式可以写为

$$\begin{aligned}\cdots+C_m\frac{dV_m}{dt}+\frac{(V_m-V_{Th})}{R_{Th}}+\frac{(V_m-V'_m)}{R_a}+C_m\frac{dV'_m}{dt}+\frac{(V'_m-V_{Th})}{R_{Th}}\\+\frac{(V'_m-V''_m)}{R_a}+C_m\frac{dV''_m}{dt}+\frac{(V''_m-V_{Th})}{R_{Th}}+\frac{(V''_m-V'''_m)}{R_a}+G_K(V'''_m-E_K)\\+G_{Na}(V'''_m-E_{Na})+\frac{(V'''_m-E_l)}{R_l}+C_m\frac{dV'''_m}{dt}+\cdots\end{aligned}\qquad(11.49)$$

除了枪乌贼巨轴突所含有的 3 种离子通道以外，神经元通常还包括其他通道，因此，神经元模型应该能够包含快速钠通道、延时钾通道以及高阈值的钙通道等其他离子通道机制。外加的离子通道可以添加到以上表达式的各个房室中，即增加以下项：

$$\sum_{i=1}^{n}G_i(V_m-E_i)$$

式中 n——各个房室的离子通道数目。

C_m、R_{Th}、R_a和 G_i的值取决于房室的大小和神经元的类型。

除了轴突以外，神经元模型可以根据需要包含任意多个树突分支，每个树突分支的模型如图 11.17 所示，其表达式为

$$\cdots+C_m\frac{dV_m}{dt}+\frac{(V_m-V_{Th})}{R_{Th}}+\frac{(V_m-V'_m)}{R_a}+C_m\frac{dV'_m}{dt}+\frac{(V'_m-V_{Th})}{R_{Th}}+\frac{(V'_m-V''_m)}{R_a}+\cdots\qquad(11.50)$$

另外，被动胞体的表达式为

$$C_m\frac{dV_m}{dt}+\frac{(V_m-V_{Th})}{R_{Th}}+\frac{(V_m-V'_m)}{R_a}\qquad(11.51)$$

主动胞体的表达式为

$$G_K(V'''_m-E_K)+G_{Na}(V'''_m-E_{Na})+\frac{(V'''_m-E_l)}{R_l}+C_m\frac{dV'''_m}{dt}\qquad(11.52)$$

以上式（11.49）～式（11.52）就是 Rodrigues 和 Enderle 在 2004 年提出的神经元整体模型。除了处于模型末端的房室以外，其他树突的房室都有两个输入，分别是前一个房室和后一个房室定义的膜电压。利用这种基本神经元模型可以生成许多神经元，并在神经元之间建立相互作用，也就是通过一个神经元突触前末梢对另一个神经元的突触后膜产生电流刺激。

图 11.34 所示的 SIMULINK 模型就模拟了两个相连神经元之间的相互作用，其

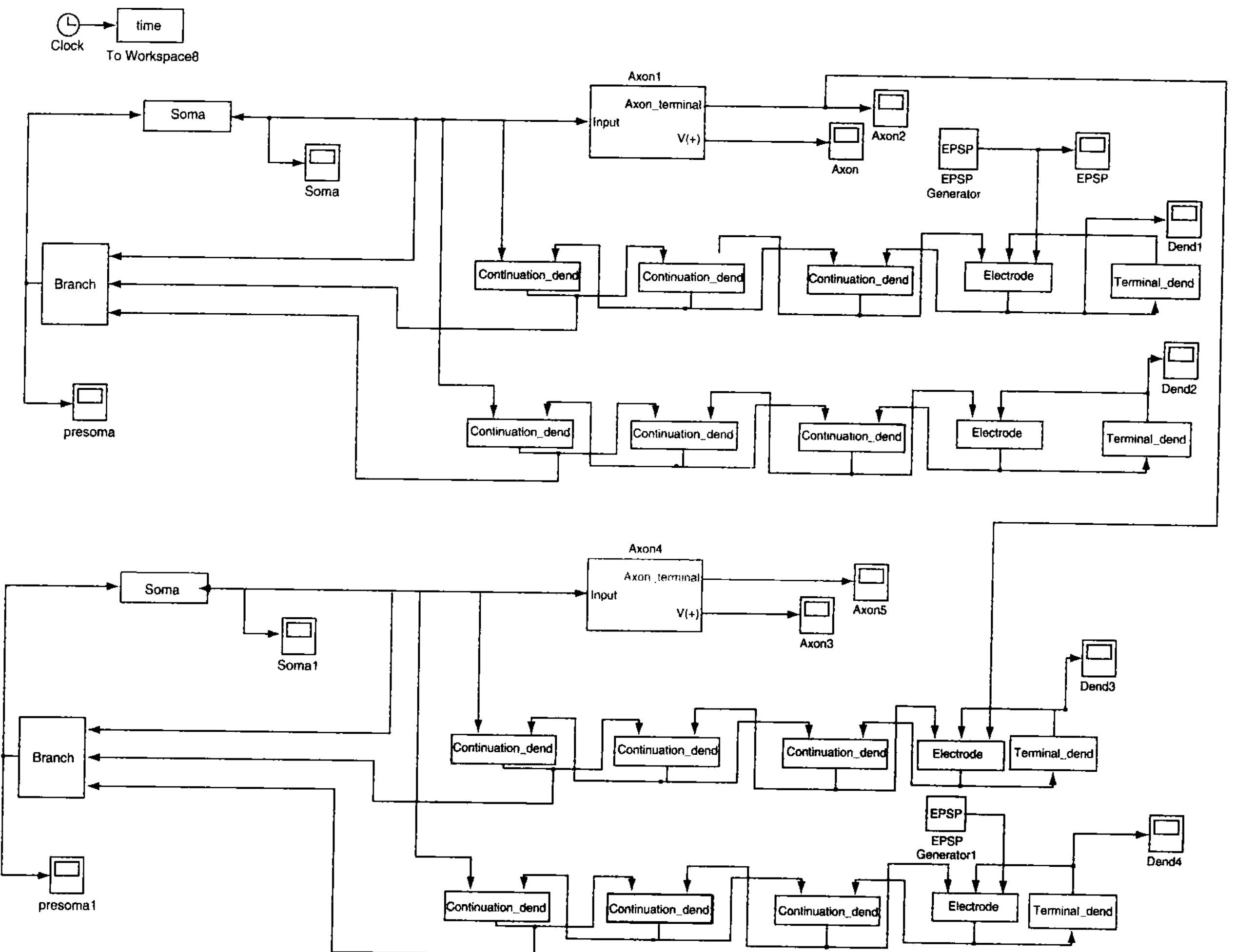
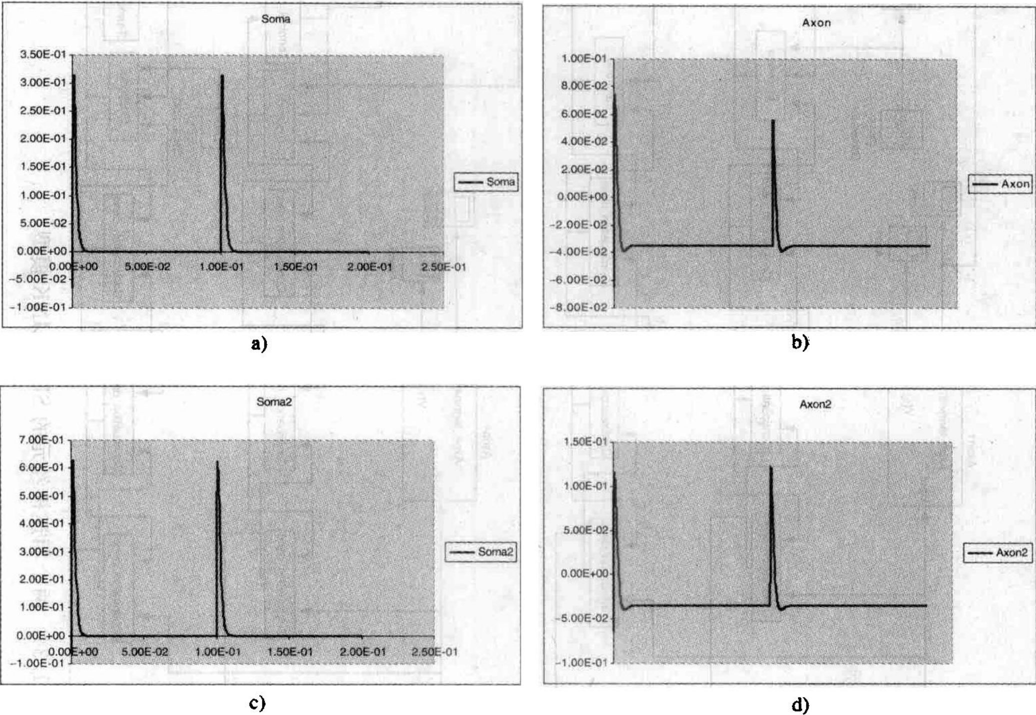


图 11.34 两个相连神经元的 SIMULINK 模型

运行结果如图 11.35 所示。该模型的轴突包含 Na^+、K^+ 和 Ca^{2+} 3 种电压门控离子通道和一种漏电流通道，每个郎飞氏节之间的有髓鞘轴突部分用 4 个被动房室来表示，整个轴突由 3 个主动房室和 2 个有髓鞘段组成。树突则由 5 个被动房室组成，而胞体则是一个被动的球形房室。外加刺激施加在第一个神经元的树突末端，该刺激用一个主动的电极房室来表示。各个轴突房室的大小是一样的，但与树突房室的大小不同。加在第一个神经元上的输入刺激信号如图 11.36 所示。



a) b) c) d)

图 11.35 图 11.34 所示模型各部分的输出信号

a）第一个神经元的胞体 b）第一个神经元的轴突 c）第二个神经元的胞体 d）第二个神经元的轴突

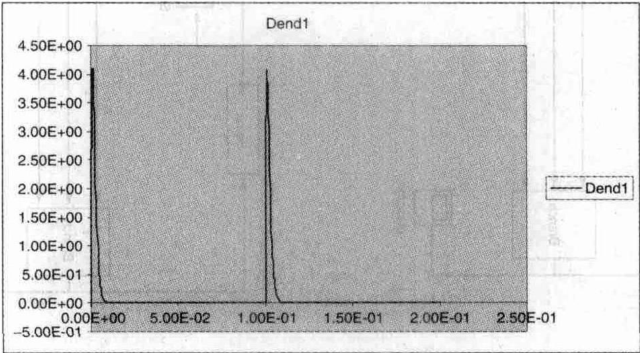


图 11.36 第一个神经元的刺激输入

虽然本章关注的是神经元，但是必须注意，很多其他细胞也会产生动作电位，并参与信号传导或者用于事件的触发，因此，本章所介绍的很多原理也同样可以用于其他细胞，只是各种动作电位的数学方程有所不同。例如，心脏产生的动作电位就要用 DiFranceso-Noble、Luo-Rudy 等模型来描述，不能用神经元的 Hodgkin-Huxley 模型。

习 题

1. 假设细胞膜只对 Ca^{2+} 离子通透，(a) 试推导 Ca^{2+} 流量的表达式；(b) 求出 Ca^{2+} 的 Nernst 电位。

2. 假设细胞膜对 Ca^{2+} 离子和 Cl^- 离子都通透，但对较大的阳离子 R^+ 不通透。并且胞内浓度为 [RCl] =100mmol/L，[$CaCl_2$]=200mmol/L；胞外浓度为 [$CaCl_2$]=300mmol/L。(a) 试推导该系统的 Donnan 平衡方程；(b) 求出 Ca^{2+} 离子的稳态平衡浓度。

3. 假设细胞膜对 Ca^{2+} 离子和 Cl^- 离子都通透，并且细胞内外溶液中都只包含这两种离子，初始时胞内浓度与胞外浓度不同。(a) 请写出 Ca^{2+} 离子流量 J_{Ca} 和 Cl^- 离子流量 J_{Cl} 的方程；(b) 试写出描述 J_{Ca} 与 J_{Cl} 之间关系的表达式；(c) 求平衡电位；(d) 求达到平衡之前跨膜电压与 [$CaCl_2$] 浓度之间的关系式。

4. 假设细胞膜只对 R^{3+} 离子通透，且胞内浓度为 [RCl_3]=2mmol/L，胞外浓度为 [RCl_3]=1.4mmol/L。(a) 试写出 R^{3+} 离子流量的表达式；(b) 推导平衡时 R^{3+} 离子的 Nernst 电位。

5. 试推导只对 Na^+、K^+ 和 Cl^- 离子通透的细胞膜的 Goldman 方程。

6. 试计算室温下蛙骨骼肌细胞的膜电压 V_m。

7. 以下是某细胞膜内外的稳态离子浓度及其渗透性，(a) 试计算 K^+ 的 Nernst 电位；(b) 应用 Goldman 方程计算静息电位；(c) 请解释空间电荷平衡原则是否满足；(d) 请解释平衡膜电位为什么不等于 0。

离 子	胞内浓度/(mmol/L)	胞外浓度/(mmol/L)	相对通透率
K^+	140	2.5	1.0
Na^+	13	110	0.019
Cl^-	3	90	0.381

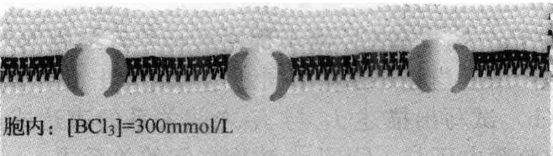
8. 假设细胞膜内外的离子浓度和通透性如下，并且室温下膜静息电位为−52mV，试计算胞内 K^+ 离子浓度。

离 子	胞内浓度/(mmol/L)	胞外浓度/(mmol/L)	相对通透率
K^+	?	4	?
Na^+	41	276	0.017
Cl^-	52	340	0.412

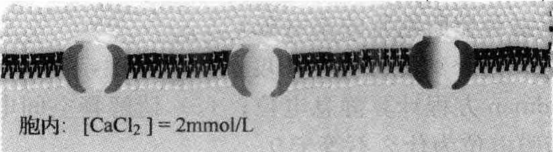
9. 假设稳态时细胞膜内外的离子浓度和通透性如下（注意 A^+ 不通透）。(a) 试计算 Cl^- 的 Nernst 电位；(b) 用 Goldman 方程计算静息电位；(c) 请解释空间电荷平衡原则是否满足；(d) 请解释平衡膜电位为什么不等于 0；(e) 请解释为什么静息电位不等于任何一种离子的 Nernst 电位。

离　子	胞内浓度/(mmol/L)	胞外浓度/(mmol/L)	相对通透率
K^+	136	15	1.0
Na^+	19	155	0.019
Cl^-	78	112	0.381
A^+	64	12	—

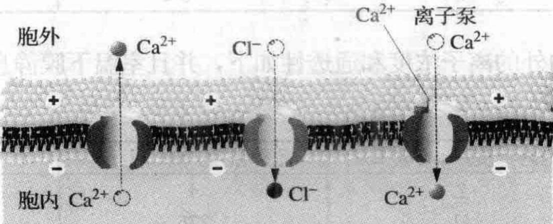
10. 假设细胞膜对 B^{3+} 和 Cl^- 离子通透，但是对较大的阳离子 R^+ 不通透，细胞膜内外的初始浓度如下图所示。(a) 试推导 Donnan 平衡方程；(b) 计算 B^{3+} 的稳态平衡浓度。



11. 下图所示的细胞膜对 Ca^{2+} 和 Cl^- 离子通透。(a) 试写出 Ca^{2+} 流量 J_{Ca} 和 Cl^- 流量 J_{Cl} 的表达式；(b) 试写出描述 J_{Ca} 与 J_{Cl} 之间关系的表达式；(c) 求平衡电位；(d) 求达到平衡之前跨膜电压与 [$CaCl_2$] 浓度之间的关系式。

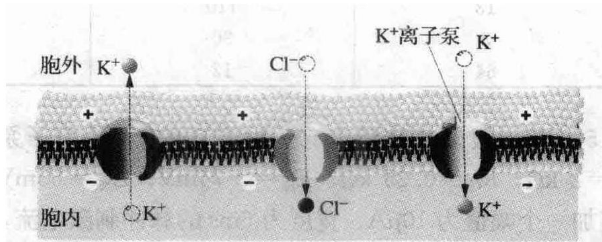


12. 下图所示细胞膜含有主动 Ca^{2+} 离子泵。假设细胞膜对 Ca^{2+} 和 Cl^- 都通透，并且 Ca^{2+} 泵的流量为 J_p，膜的厚度设为δ。试写出 Ca^{2+} 泵流量随 [Ca^{2+}] 浓度变化的函数。

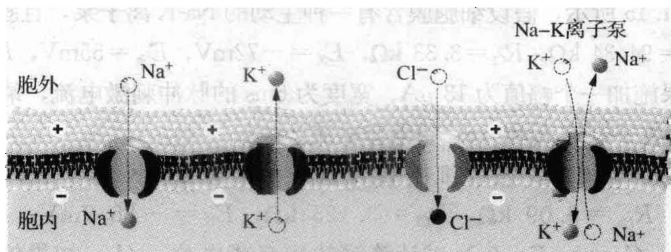


13. 下图所示细胞膜对 K^+ 和 Cl^- 离子通透，并且主动离子泵将 K^+ 从胞外泵入胞内，膜的厚度设为 δ。(a) 写出各离子流量的表达式；(b) 计算平衡时的离子流量；

(c) 写出离子泵流量随浓度差 ($[K^+]_i-[K^+]_o$) 变化的函数；(d) 定性描述细胞膜两侧的离子浓度。



14. 下图所示的细胞膜包含 2 种离子的主动泵。假设细胞膜对 Na^+、K^+ 和 Cl^- 离子通透，并且存在流量关系 $J_p(K)=J_p(Na)=J_p$，膜的厚度设为δ。试写出浓度差 ($[Cl^-]_i-[Cl^-]_o$) 随流量 J_p 变化的函数。



15. 假设稳态时细胞膜内外的离子浓度和渗透性如下 (注意 A^+ 不通透)，并且各离子通道的电阻为 $R_K=1.7\ k\Omega$，$R_{Na}=9.09\ k\Omega$，$R_{Cl}=3.125\ k\Omega$。(a) 试计算各种离子的 Nernst 电位；(b) 画出该细胞膜的电路模型 (提示：参照图 11.13)；(c) 利用 (b) 中所建立的电路模型计算膜静息电位；(d) 求 (b) 中所建电路的戴维南等效电路。

离　子	胞内浓度/(mmol/L)	胞外浓度/(mmol/L)	相对通透率
K^+	168	6	1.0
Na^+	50	337	0.019
Cl^-	41	340	0.381
A^+	64	12	—

16. 假设图 11.13 所示细胞膜的参数为 $R_K=0.1\ k\Omega$，$R_{Na}=2\ k\Omega$，$R_{Cl}=0.25\ k\Omega$，$E_K=-74mV$，$E_{Na}=55mV$，$E_{Cl}=-68mV$。(a) 请计算膜电压 V_m；(b) 求出该电路模型的戴维宁等效电路。

17. 如图 11.14 所示，假设细胞膜含有一种主动的 Na-K 离子泵，且膜的参数为 $R_K=0.1\ k\Omega$，$R_{Na}=2\ k\Omega$，$R_{Cl}=0.25\ k\Omega$，$E_K=-74mV$，$E_{Na}=55mV$，$E_{Cl}=-68mV$。请计算主动离子泵的电流 I_{Na} 和 I_K。

18. 假设稳态时细胞膜内外的离子浓度和渗透性如下表 (注意 A^+ 不通透)。(a) 如果 $R_K=1.7\ k\Omega$ 且 $R_{Cl}=3.125\ k\Omega$，请计算 R_{Na}。(b) 求戴维南等效电路模型。

离　子	胞内浓度/(mmol/L)	胞外浓度/(mmol/L)	相对通透率
K^+	140	2.5	1.0
Na^+	13	110	0.019
Cl^-	3	90	0.381
A^+	64	12	—

19. 如图 11.15 所示，假设细胞膜含有一种主动的 Na-K 离子泵，且膜的参数为 $R_K=0.1\ k\Omega$，$R_{Na}=2\ k\Omega$，$R_{Cl}=0.25\ k\Omega$，$E_K=-74mV$，$E_{Na}=55mV$，$E_{Cl}=-68mV$，$C_m=1\ \mu F$。如果施加一个幅值为 $10\mu A$、宽度为 6ms 的脉冲刺激电流，(a) 请计算膜电压 V_m；(b) 计算膜电容电流；(c) 如果刺激电流的脉冲宽度从 6ms 减到 1ms，计算要使膜电压 V_m 升至 $-40mV$ 所需的脉冲电流幅值；(d) 如果阈值电压为 $-40mV$，施加的刺激是 (c) 中所求得的脉冲电流，请解释是否会产生动作电位。

20. 如图 11.15 所示，假设细胞膜含有一种主动的 Na-K 离子泵，且膜的参数为 $R_K=2.727\ k\Omega$，$R_{Na}=94.34\ k\Omega$，$R_{Cl}=3.33\ k\Omega$，$E_K=-72mV$，$E_{Na}=55mV$，$E_{Cl}=-49.5mV$，$C_m=1\ \mu F$。如果施加一个幅值为 $13\ \mu A$、宽度为 6ms 的脉冲刺激电流，请计算 (a) 膜电压 V_m；(b) K^+ 离子电流 I_K 和 (c) 膜电容电流。

21. 如图 11.15 所示，假设细胞膜含有一种主动 Na-K 离子泵，且膜的参数为 $R_K=1.75\ k\Omega$，$R_{Na}=9.09\ k\Omega$，$R_{Cl}=3.125\ k\Omega$，$E_K=-85.9mV$，$E_{Na}=54.6mV$，$E_{Cl}=-9.4mV$，$C_m=1\ \mu F$。(a) 试计算膜的静息膜电位；(b) 如果施加较小的阈下电流脉冲刺激，试求解膜电压 V_m 的变化。

22. 如图 11.15 所示，假设细胞膜含有一种主动 Na-K 离子泵，且膜的参数为 $R_K=2.727\ k\Omega$，$R_{Na}=94.34\ k\Omega$，$R_{Cl}=3.33\ k\Omega$，$E_K=-72mV$，$E_{Na}=55mV$，$E_{Cl}=-49.5mV$，$C_m=1\ \mu F$。请设计一种刺激电流脉冲，可以在 3ms 内驱使膜电压 V_m 达到阈值。假设阈值电压为 $-40mV$。(a) 求电流脉冲的幅值和宽度；(b) 求 V_m 并作图显示。

23. 假设有 $20\mu A$ 的电流脉冲流过枪乌贼巨轴突的细胞膜，且轴突的 Hodgkin-Huxley 参数为 $G_l=\frac{1}{R_l}=0.3\times10^{-3}\ S$，$\overline{G}_K=36\times10^{-3}\ S$，$\overline{G}_{Na}=120\times10^{-3}\ S$，$E_K=-72\times10^{-3}V$，$E_l=-49.4\times10^{-3}V$，$E_{Na}=55\times10^{-3}V$，$C_m=1\times10^{-6}F$，请仿真所产生的动作电位，并作出 (a) V_m、G_{Na} 和 G_K 随时间变化的曲线；(b) V_m、n、m 和 h 随时间变化的曲线；(c) V_m、I_{Na}、I_K、I_C 和 I_l 随时间变化的曲线。

24. 假设有 $20\mu A$ 的电流脉冲流过轴突细胞膜，轴突的参数为 $G_l=\frac{1}{R_l}=0.3\times10^{-3}S$，$\overline{G}_K=36\times10^{-3}S$，$\overline{G}_{Na}=120\times10^{-3}S$，$E_K=-12\times10^{-3}V$，$E_l=10.6\times10^{-3}V$，$E_{Na}=115\times10^{-3}V$，$C_m=1\times10^{-6}F$。假设用式 (11.41) ～式 (11.48) 描述轴突，请仿真所产生的动作电位，并作出 (a) V_m、G_{Na} 和 G_K 随时间变化的曲线；(b) V_m、n、m 和 h 随时间变化的曲线；(c) V_m、I_{Na}、I_K、I_C 和 I_l 随时间变化的曲线。

25. 本习题用于考察枪乌贼巨轴突中阈值电位对于动作电位的影响。轴突的

Hodgkin-Huxley 参数为 $G_l=\frac{1}{R_l}=0.3\times10^{-3}\,\text{S}$，$\overline{G}_K=36\times10^{-3}\,\text{S}$，$\overline{G}_{Na}=120\times10^{-3}\,\text{S}$，$E_K=-72\times10^{-3}\,\text{V}$，$E_l=-49.4\times10^{-3}\,\text{V}$，$E_{Na}=55\times10^{-3}\,\text{V}$，$C_m=1\times10^{-6}\,\text{F}$。(a) 假设给枪乌贼巨轴突细胞膜施加一个幅值为 $-10\mu\text{A}$ 且持续时间很长的电流脉冲，使细胞膜超极化；然后，在 $t=0$ 时刻，将电流脉冲撤除。请仿真由此产生的动作电位。(b) 阈值电位被定义为 $I_{Na}>I_K+I_l$时的电位，将 α_m方程中的数值 25 减小就可以方便地改变阈值电位。假设 α_m方程中的 25 改为 10，并且用幅值为 $-10\mu\text{A}$ 的长时间电流脉冲使细胞膜超极化，并在 $t=0$ 时刻撤除电流脉冲。请仿真由此产生的动作电位。

26. 请仿真计算图 11.26 所示电压钳实验的曲线。

27. 请设计一种输入刺激电流的波形，用于研究动作电位的不应期。(提示：用双脉冲电流刺激信号。) 最小的不应期有多长？如果在最小不应期结束之前施加了第二个刺激脉冲，那么，刺激幅值要多大才能诱发出动作电位？

28. 请解释下图所示的电路是否能够完成电压钳实验。图示的无髓鞘轴突两端都被封住，且轴突中插入了一根银丝用于消除胞内电阻 R_a。要求说明你所作的任何假设。

推荐阅读资料

Bahill, A.T. (1981). *Bioengineering: Biomedical, Medical and Clinical Engineering*. Prentice-Hall, Englewood Cliffs, NJ.

Bronzino, J.D. (2000). *The Biomedical Engineering Handbook*. CRC, Boca Raton, FL.

Bower, J.M. and Beeman, D. (1998). *The Book of Genesis: Exploring Realistic Neural Models with the General Neural Simulation System*. Springer-Verlag, New York.

Deutsch, S. and Deutsch, A. (1993). *Understanding the Nervous System—An Engineering Perspective*. IEEE Press, New York.

DiFrancesco, D. and Noble, D. (1985). A model of cardiac electrical activity incorporating ionic pumps and concentration changes. *Phil Trans R. Soc London [B]* **307**, 307–353.

Enderle, J.D. (2002). Neural control of saccades. In J. Hyönä, D. Munoz, W. Heide, and R. Radach (Eds.), *The Brain's Eyes: Neurobiological and Clinical Aspects to Oculomotor Research, Progress in Brain Research, V. 140*. Elsevier, Amsterdam.

Fulton, J.F. and Cushing, H. (1936). A bibliographical study of the Galvani and Aldini writings on animal electricity. *Ann. Sci.* **1**, 239–268.

Guyton, A.C. and Hall, J. E. (1995). *Textbook on Medical Physiology*, 9th Ed. Saunders, Philadelphia.

Hille, B. (1992). *Ionic Channels of Excitable Membranes*, 2nd Ed. Sunderland, Massachusetts.

Hodgkin, A., Huxley, A. and Katz, B. (1952). Measurement of current-voltage relations in the membrane of the giant axon of *Loligo*. *J. Physiol. (London)* **116**, 424–448.

Hodgkin, A. and Huxley, A. (1952a). Currents carried by sodium and potassium ions through the membrane of the giant axon of *Loligo*. *J. Physiol. (London)* **116**, 449–472.

Hodgkin, A. and Huxley, A. (1952b). Currents carried by sodium and potassium ions through the membrane of the giant axon of *Loligo*. *J. Physiol. (london)* **116**, 49–472

Hodgkin, A. and Huxley, A. (1952c). The components of membrane conductance in the giant axon of *Loligo*. *J. Physiol. (London)* **116**, 473–496.

Hodgkin, A. and Huxley, A. (1952d). The dual effect of membrane potential on sodium conductance in the giant axon of *Loligo*. *J. Physiol. (London)* **116**, 497–506.

Hodgkin, A. and Huxley, A. (1952e). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol. (London)* **117**, 500–544.

Kandel, E.R., Schwartz, J.H. and Jessell, T.M. (2000). *Principles of Neural Science*, 5th Ed. McGraw-Hill, New York.

Keener, J. and Sneyd, J. (1998). *Mathematical Physiology*. Springer, New York.

Luo, C. and Rudy, Y. (1994). A dynamic model of the cardiac ventricular action potential: I. Simulations of ionic currents and concentration changes. *Circ. Res.* **74**, 1071.

Matthews, G.G. (1991). *Cellular Physiology of Nerve and Muscle*. Blackwell Scientific, Boston.

Nernst, W. (1889). Die elektromotorishe Wirksamkeit der Jonen. *Z. Physik. Chem.* **4**, 129–188.

Northrop, R. (2001). *Introduction to Dynamic Modeling of Neurosensory Systems*. CRC, Boca Raton.

Plonsey, R. and Barr, R.C. (1982). *Bioelectricity: A Quantitative Approach*. Plenum, New York.

Rinzel, J. (1990). Electrical excitability of cells, theory and experiment: Review of the Hodgkin-Huxley foundation and an update. *Bull. Math. Biology: Classics of Theoretical Biology* **52**, 5–23.

Rodriguez Campos, F. and Enderle, J.D. (2004). Porting Genesis to SIMULINK, *Proceedings of the 30th IEEE EMBS Annual International Conference*, September 2–5.

第 12 章　生理系统仿真建模

John Enderle 博士[⊖]

本章目录

⊖ R. J. Fisher 也参与了本章部分内容的撰写。——原文注

本章主要内容:

- 生理系统数学模型的建立过程;
- 房室模型的基本概念;
- 应用房室模型分析生理系统;
- 非线性房室模型的求解;
- 人眼扫视活动的定性分析;
- 扫视眼动系统的二阶模型;
- 扫视眼动系统的脉冲一阶跃输入控制信号;
- 应用非线性和线性肌肉模型分析眼肌的作用机制;
- 扫视眼动系统的四阶仿真模型;
- 应用系统辨识方法估计模型参数。

12.1 绪 论

定量生理学模型就是仿真实际生理系统行为和特性的数学方程。生物学家一般不用数学方程,他们习惯于定性描述实际的生理系统。但是,本章将介绍的定量生理学模型要比定性描述更有用。生理系统几乎都是动态变化的系统,在数学上要用微分方程来描述;并且,本章所述的建模方法与生理学、生物物理学和生物化学等许多其他学科密切相关,交叉在一起,并包括各种电学和力学的仿真模拟。数学模型通常根据基本物理定律建立,本章将介绍几种较复杂的、较大规模的生理系统模型。

建立模型之前总是要做实验,并采集数据。实验最好能够提供与模型中所用变量相关的数据,因此,实验的设计和完成是建模过程中最重要且最耗时的工作。根据基本物理定律建立好模型之后,它就成了一种工具,可用于解释实验数据所隐含的机制,并且预测实验没能采集的数据,也就是,预测其他输入刺激条件下系统可能产生的响应。模型是分析和整理复杂的实验数据并验证假说的工具,建立模型最重要的目标是形成新知识,预测尚未发生的现象,并辅助新实验的设计。

图 12.1 说明了建立模型的基本步骤。首先考察实验现象,给出生理系统的定性描述和推测,并用数学方程建立一个初步的假设模型。然后,用采集的实验数据测试模型的可靠性。如果模型的输出与实验数据之间吻合很好,测试结果令人满意,那么这个数学模型就完成了。但是,如果模型没有达到预期的目标,那么就需要补充实验数据,再修正模型。模型中的变量并不都是可以通过实验测得的,某些变量可以测量,而某些变量却不能测量。新补充的实验可以进一步提供某些变量的信息,加深我们对于生理系统的了解,以便对模型进行改进。这样,利用实验数据反复测

试模型，直到令人满意为止。通常要用统计学方法检验模型与实验数据之间的吻合程度，好的模型必须能够正确预测生理系统的反应。

数字计算机、程序设计语言和各种仿真软件的出现使生理模型的应用迅速发展。以前，没有计算机的时候，生物医学系统的数学模型要么过于简单，要么像 Hodgkin 和 Huxley 在 1952 年发表的论文中所述的那样，手工计算的工作量非常大。如今，计算机已相当普及，以至于“建模”和“计算机仿真”几乎成了同义词。使得所建立的模型可以尽可能多地包含有关生理系统结构及其内部相互关系的信息，更接近真实系统，无需顾及任何计算量的问题。虽然模型的设计确实可以变得越来越复杂，但必须正确估计每一步复杂化所换来的价值。也就是，一方面，在确保模型具有实际意义的前提下，要使模型能够尽可能简单地解释实验数据，但不能简单得毫无意义；另一方面，过于复杂的模型也没什么用处。

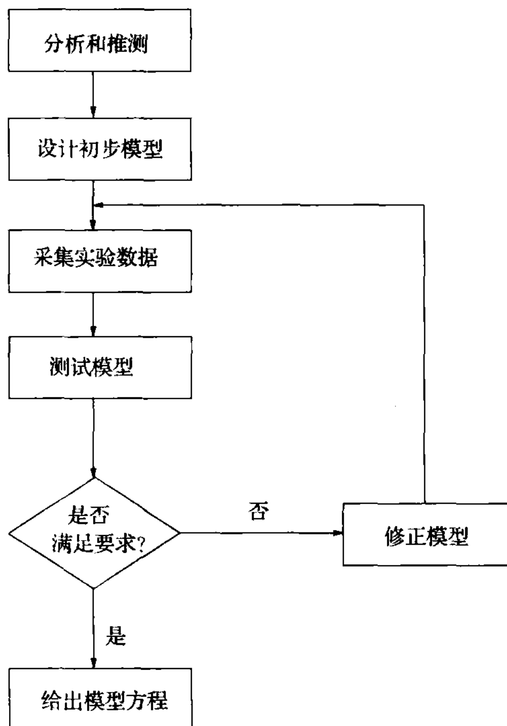


图 12.1　建立数学模型的流程图

12.1.1　确定性模型和随机性模型

确定性模型的自变量之间，以及自变量与因变量之间具有明确的关系。对于同样一组给定的初始条件，这种模型每次运行都会产生完全相同的结果。而随机性模型包含随时间等因素变化的随机变量。对于同样一组给定的初始条件，这种模型每次运行产生的结果有所不同。可以肯定，实际模型中随机性模型远比确定性模型多得多。

注意，所有确定性模型都含有测量误差，测量误差给确定性模型带来了随机因素，于是，这种模型也可以看作具有随机性。不过，本章将主要特性为确定性因素的模型都看作确定性模型，而将主要特性是随机因素的模型看作随机性模型。下面主要介绍确定性模型。

12.1.2　模型的求解方法

模型一般有两种求解方法，一种是求模型的显式解析解，可以用传统的拉普拉斯变换等解析求解微分方程的方法来求解。例如，给定如下微分方程

$$\ddot{x}+4\dot{x}+3x=9$$

设初始条件为 $x(0)=0$，$\dot{x}(0)=1$，则方程的解为

$$x(t)=-4\mathrm{e}^{-t}+\mathrm{e}^{-3t}+3$$

对于没有解析解的模型，可以求其数值解，也就是仿真解。例如，以下方程就是如此：

$$x=\int_{-20}^{20}\frac{1}{33\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}\left(\frac{t-7}{33}\right)^{2}}\mathrm{d}t$$

该方程是求高斯曲线下所包含的面积，它没有解析解，因此，必须用梯形求积公式等近似算法来求解。非线性微分方程一般都没有解析解，因此都要用迭代法或者 SIMULINK 等仿真软件包来求解，前面第 11 章所介绍的 Hodgkin-Huxley 模型就是这样求解的。

正向和反向求解

很多工程领域中，各种工程系统一般都根据事先预定的技术指标来设计和制造。由于数学模型既经济又有效，因此，工程师们经常先建立可以模拟系统的完整数学模型，来预测系统的行为和特性。所建的模型包含许多参数，通过选择各种不同的参数值，可以获得一组特定的技术指标，如上升时间、建立时间和峰值时间等。

不过，生物医学工程师建立生理系统的数学模型，并不是为了制造生理系统，而只是为了研究系统的行为，也就是系统的输入和输出。如图 12.1 所示，建立模型的步骤包括：分析和推测模型的结构形式，收集实验数据，再利用实验数据估计模型的各个参数。生理系统建模的目标并不是要设计该系统，而是要识别系统的组成部分（即各个参数）。很多时候，建模所需的数据并不能用本书第 8～10 章所介绍的现有传感器和生物仪器直接测量。一般，需要经过某种转换，才能将测量数据转变为模型所需变量的估计值。采集合适的实验数据往往是生理系统建模中最难的工作。

建立模型一般需要估计模型的参数，使得模型的输出（即预测值）$\hat{x}_i$与实际实验数据 x_i之间的误差最小。例如，估计模型参数的一种衡量指标就是求模型预测值与实验数据之间的误差平方之和 S，并使其最小化，即

$$S=\sum_{i=1}^{n}\varepsilon_i^2=\sum_{i=1}^{n}(x_i-\hat{x}_i)^2$$

式中　ε_i——模型预测值 $\hat{x}_i$与实验测量数据 x_i之间的误差。

该方法是模型预测值对于测量数据的一种无偏估计。

下面，本章第一部分将首先介绍房室模型，它是人体各种生理系统建模的通用方法。本章第二部分主要介绍一个特殊的人体系统——快速眼动系统的模型，将从 20 世纪 20 年代开始建立的早期肌肉模型，一直讲到如今的神经网络控制模型。通过这些内容的学习，可以使读者对于图 12.1 所示的建模过程有较深入的了解。快速眼动系统可能是人体所有系统中研究最深入的系统之一。其建模成功的原因有以下几点：实验数据比较容易获得；比其他生理系统简单；系统动态变化过程中不存在反

馈机制。本章最后将介绍系统辨识，即参数估计。

12.2 房室模型

人体的呼吸和循环等生理系统都涉及物质从一个房室到另一个房室的传输过程，这种系统的分析用房室模型比较方便。20 世纪 20 年代开始进行的有关化合物示踪剂的代谢研究给房室建模打下了基础。房室建模与其他生理建模的不同之处在于，它关注的是如何维持体内化学物质的正常浓度水平和流量。

人体内很容易识别的几个房室是：

1）细胞房室：由细胞膜将其与胞外空间隔离；

2）组织间隙房室：含有细胞孵育液，与血浆之间由毛细血管壁隔开；

3）血浆房室：属于人体循环系统，是浸浴各种血细胞的液体。

房室分析的变量有物质（即溶质）的质量和浓度、温度以及压力等。分析物有药物和示踪剂等外源性物质，或者葡萄糖、酶、胰岛素等内源性物质。这些物质从一个房室传输到另一个房室的过程遵循质量守恒定律。所建立的房室模型可以是线性的，也可以是非线性的。可见，房室分析是一种统一的理论，可以用于人体各种生理系统的建模仿真。本书第 11 章中已经介绍过的描述跨膜电流的神经元模型，就是房室模型的一个例子。

虽然本章所述的房室分析法主要针对人体的生理系统，但很多科学家也将这种方法用于药物代谢动力学、生化反应工程、流体输运甚至半导体设计和制造等领域的研究。在分析几个较大的生理系统之前，我们先从房室分析的角度介绍 Fick 扩散定律的推导过程。

12.2.1 由薄膜隔开的两个房室之间的物质传输

本节利用 Fick 扩散定律来研究被薄膜隔开的两个房室之间的物质传递过程。Fick 扩散定律如下

$$\frac{\mathrm{d}q}{\mathrm{d}t}=-DA\frac{\mathrm{d}c}{\mathrm{d}x} \tag{12.1}$$

式中　q——溶质质量；

A——薄膜表面积；

c——溶液浓度；

D——扩散系数；

$\mathrm{d}x$——薄膜厚度。

一般不用溶质的质量，而用溶液的浓度，质量与浓度之间的换算关系为

$$浓度=\frac{质量}{体积}$$

设系统的两个房室如图 12.2 所示，其中，V_1和 V_2分别是房室 I 和房室 II 的体积；q_1和 q_2分别是房室 I 和房室 II 的溶质质量；c_1和 c_2则分别是房室 I 和房室 II 的溶质浓度。假设初始时房室 I 中加入了质量为 Q_{10}的溶质，那么，房室 I 的溶质变化率就是

$$\frac{dq_1}{dt}=-DA\frac{dc_1}{dx}=-DA\frac{(c_1-c_2)}{\Delta x} \tag{12.2}$$

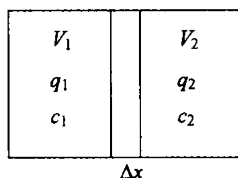


图 12.2 双房室扩散模型，其中 Δx 为薄膜厚度（单位为 cm）

再将溶质质量用浓度代替，即

$$q_1=V_1c_1 \tag{12.3}$$

此式求微分，可得

$$\frac{dq_1}{dt}=V_1\frac{dc_1}{dt} \tag{12.4}$$

将上式代入式（12.2），即得

$$V_1\frac{dc_1}{dt}=\frac{-DA}{\Delta x}(c_1-c_2) \tag{12.5}$$

其中 $K=\frac{DA}{\Delta x}$定义为传输率，将其代入式（12.5），可得

$$\frac{dc_1}{dt}=-\frac{K}{V_1}(c_1-c_2) \tag{12.6}$$

根据质量守恒定律，有

$$q_1+q_2=Q_{10}$$

转换成浓度形式为

$$V_1c_1+V_2c_2=V_1C_{10} \tag{12.7}$$

式中 C_{10}——起始溶质加入时房室 I 的初始浓度。

由式（12.7）可得房室 II 的浓度为

$$c_2=\frac{C_{10}V_1-V_1c_1}{V_2} \tag{12.8}$$

将其代入式（12.6），得到

$$\frac{dc_1}{dt}=\frac{-K}{V_1V_2}[V_2c_1-V_1C_{10}+V_1c_1]=\frac{KC_{10}}{V_2}-\frac{Kc_1}{V_1V_2}(V_1+V_2)$$

即

$$\frac{dc_1}{dt}+K\left(\frac{V_1+V_2}{V_1V_2}\right)c_1=\frac{KC_{10}}{V_2} \tag{12.9}$$

这是一阶线性微分方程，其强制函数为

$$f(t)=\frac{KC_{10}}{V_2} \tag{12.10}$$

且初始条件为 $c_1(0)=C_{10}$，用标准方法可以求解该线性微分方程。

为了简化，假设 $V_1=V_2$，于是，式（12.9）变成

$$\frac{dc_1}{dt}+\frac{2K}{V_1}c_1=\frac{KC_{10}}{V_1} \tag{12.11}$$

注意，此方程的根为$-\frac{2K}{V_1}$，因此，其自由解为

$$c_{1_n}=B_1 e^{-\frac{2Kt}{V_1}} \tag{12.12}$$

式中　B_1——由初始条件决定的常数。

方程的强迫响应与式（12.10）的强制函数具有相同的形式，将$c_{1_f}=B_2$代入式（12.11），可得

$$\frac{2K}{V_1}B_2=\frac{KC_{10}}{V_1}$$

即$B_2=C_{10}/2$。于是，全响应为

$$c_1=c_{1_n}+c_{1_f}=B_1 e^{-\frac{2Kt}{V_1}}+\frac{C_{10}}{2} \tag{12.13}$$

下面用初始条件求B_1，即

$$c_1(0)=C_{10}=B_1 e^{-\frac{2Kt}{V_1}}\bigg|_{t=0}+\frac{C_{10}}{2}=B_1+\frac{C_{10}}{2}$$

可得$B_1=C_{10}/2$。于是，$t\geqslant 0$时，全响应为

$$c_1=\frac{C_{10}}{2}\left(e^{-\frac{2Kt}{V_1}}+1\right)$$

注意，利用式（12.8）可以求得房室 II 的浓度为

$$c_2=\frac{V_1C_{10}-V_1c_1}{V_2}=\frac{C_{10}}{2}\left(1-e^{-\frac{2Kt}{V_1}}\right)$$

这种扩散模型是双房室模型的一个例子，可以用于描述细胞膜的跨膜扩散活动，此时dx即表示细胞膜厚度。

12.2.2　建立房室模型的基本方法

房室建模就是用有限个房室描述某个系统，每个房室都有物质的进出，房室之间通过扩散进行物质交换，如放射性示踪剂、葡萄糖或者胰岛素分子、氧气或者二氧化碳气体等物质的交换都是如此。房室模型利用质量守恒定律（即变化量等于输入量减去输出量）来预测各个房室中物质浓度随时间的变化过程。模型可以是线性的、非线性的、连续的、离散的甚至含有时变参数或随机参数。如果模型是线性连续的，那么，物质浓度的变化可以用各个指数项之和来表示。模型的示意图中用方框表示房室，用箭头表示房室之间的物质流动。

描述两个房室之间物质的扩散传输时，有如下假设：

1）每个房室的容积保持不变，为常数；

2）进入房室的任何物质q都瞬间就与整个房室的溶液混合在一起，均匀分布；

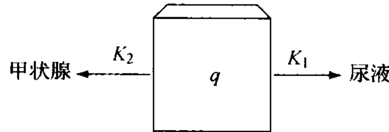
3）房室中物质的流失速率正比于其物质质量与传输率K的乘积（即Kq）。

建立模型时，较困难的一步是识别整个系统所含有的房室，并确定其数目。由于某些房室的数据根本无法直接测量，因此，采集相关的测量数据也是很难的一步。这两步工作都不是本书所要介绍的内容，有兴趣的读者可以参阅本章末尾所列的相

关参考书。

最简单的房室模型只有一个房室，这种情况可以看作双房室模型的一个特例，也就是将离开单房室进入其他空间的物质看作进入另一个房室即可。

例 12.1 下图所示是血浆单房室模型的示意图，其中 q 为房室中放射性碘同位素 I^{131} 的质量，K_1 和 K_2 为传输率。



I^{131} 通过消化系统被人体吸收之后，就立即进入血浆；然后，又从血浆进入甲状腺，并随尿液排出体外。假设 I^{131} 的初始质量为 $q(0)$，请建立该房室模型，并计算一次性服用 I^{131} 示踪剂之后，血浆房室的 I^{131} 响应过程。

解：根据质量守恒定律，即变化量＝输入量－输出量，可以建立描述房室中 I^{131} 质量变化率的微分方程，即

$$\text{变化量}=\frac{\mathrm{d}q}{\mathrm{d}t}$$

$$\text{输入量}=0$$

$$\text{输出量}=K_1q+K_2q$$

于是

$$\frac{\mathrm{d}q}{\mathrm{d}t}=-(K_1+K_2)q$$

该微分方程的解与前一节所述方程的解相似，$t \geqslant 0$ 时有

$$q=q(0)\mathrm{e}^{-(K_1+K_2)t}$$ ■

本例所述方法在医学诊断中有许多应用，稍作修改之后，其应用范围更广。此例假设物质进入体内就能够立即均匀分布于血浆房室之中，其实在生理学上这是不可能的，任何物质都有吸收过程。不过，在数学上可以实现，将输入设置为冲激函数 $\delta(t)$，就可以瞬间改变系统的初始状态。医学实践中最接近冲激函数的输入是快速静脉注射，被称为“团注”。

如图 12.3 所示是一个更接近实际情况的双房室模型。物质先由消化系统吸收，然后通过新陈代谢从尿液排出。模型中添加消化系统这个房室之后，物质不再瞬间直接进入血浆，而是逐渐从消化系统进入血浆。根据质量守恒定律，可以写出两个房室的微分方程

$$\frac{\mathrm{d}q_1}{\mathrm{d}t}=K_3q_2-(K_1+K_2)q_1$$

$$\frac{\mathrm{d}q_2}{\mathrm{d}t}=-K_3q_2$$

第二个方程只含变量 q_2，其解是 $q_2=q_2(0)\ \mathrm{e}^{-K_3t}$，其中 $q_2(0)$ 表示被吸收物质的质量。将此 q_2 的解代入第一个方程，可得

$$\frac{dq_1}{dt}=q_2(0)K_3e^{-K_3t}-(K_1+K_2)q_1$$

重排此式，得到

$$\frac{dq_1}{dt}+(K_1+K_2)q_1=q_2(0)K_3e^{-K_3t} \tag{12.14}$$

这是一阶微分方程，其强制函数为 $q_2(0)K_3e^{-K_3t}$，自由解是 $q_{1_n}=B_1e^{-(K_1+K_2)t}$，强迫响应为 $q_{1_f}=B_2e^{-K_3t}$，因此，全解如下

$$q_1=q_{1_n}+q_{1_f}=B_1e^{-(K_1+K_2)t}+B_2e^{-K_3t} \tag{12.15}$$

将强迫响应代入式（12.14），可以求 B_2，即

$$-K_3B_2e^{-K_3t}+(K_1+K_2)B_2e^{-K_3t}=q_2(0)K_3e^{-K_3t}$$

求解此式，得 B_2

$$B_2=\frac{q_2(0)K_3}{K_1+K_2-K_3}$$

至此，全响应为

$$q_1=B_1e^{-(K_1+K_2)t}+\frac{q_2(0)K_3}{K_1+K_2-K_3}e^{-K_3t} \tag{12.16}$$

利用初始条件 $q_1(0)=0$，可以求得 B_1，即

$$q_1(0)=0=\left[B_1e^{-(K_1+K_2)t}+\frac{q_2(0)K_3}{K_1+K_2-K_3}e^{-K_3t}\right]_{t=0}=B_1+\frac{q_2(0)K_3}{K_1+K_2-K_3}$$

可得

$$B_1=-\frac{q_2(0)K_3}{K_1+K_2-K_3}$$

且

$$q_1=\frac{q_2(0)K_3}{K_1+K_2-K_3}\left(e^{-K_3t}-e^{-(K_1+K_2)t}\right) \tag{12.17}$$

用浓度表示，$t\geqslant 0$ 时的解就是

$$c_1=\frac{1}{V_1}\frac{q_2(0)K_3}{(K_1+K_2-K_3)}\left(e^{-K_3t}-e^{-(K_1+K_2)t}\right) \tag{12.18}$$

为了计算血浆房室Ⅰ中物质含量达到最大值时的时刻，求式（12.17）对时间 t 的微分，并令其微分值等于 0，即：

$$\begin{aligned}\frac{dq_1}{dt}&=\frac{d}{dt}\left(\frac{q_2(0)K_3}{K_1+K_2-K_3}(e^{-K_3t}-e^{-(K_1+K_2)t})\right)\\&=\frac{q_2(0)K_3}{K_1+K_2-K_3}\left(-K_3e^{-K_3t}+(K_1+K_2)e^{-(K_1+K_2)t}\right)\end{aligned} \tag{12.19}$$

将此式置 0，得

$$\frac{q_2(0)K_3}{K_1+K_2-K_3}\left(-K_3e^{-K_3t}+(K_1+K_2)e^{-(K_1+K_2)t}\right)=0$$

即

$$K_3e^{-K_3t}=(K_1+K_2)e^{-(K_1+K_2)t}$$

两边同乘以 $e^{(K_1+K_2)t}$，并除以 K_3，可得

$$e^{(K_1+K_2)t}e^{-K_3t}=e^{(K_1+K_2-K_3)t}=\frac{K_1+K_2}{K_3}$$

再将其两边取对数

$$(K_1+K_2-K_3)t=\ln\left(\frac{K_1+K_2}{K_3}\right)$$

于是，求得 t_{max} 为

$$t_{max}=\frac{\ln\left(\frac{K_1+K_2}{K_3}\right)}{(K_1+K_2-K_3)} \qquad (12.20)$$

由此式可见，(K_1+K_2) 比 K_3 小得越多，血浆中物质含量达到最大值所需的时间就越长。

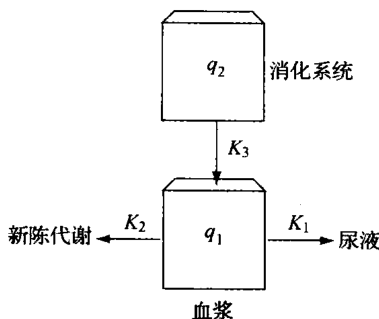


图 12.3　包含实际消化系统的双房室模型，血浆中的物质通过新陈代谢和尿液排出

例 12.2　设图 12.3 所示房室模型中，$K_1+K_2=0.005\text{min}^{-1}$，$K_3=0.02\text{min}^{-1}$，且被吸收的物质为 50g。请计算血浆中该物质能够达到的最大含量。

解：由式（12.20）可得

$$t_{max}=\frac{\ln\left(\frac{K_1+K_2}{K_3}\right)}{(K_1+K_2-K_3)}=\frac{\ln\left(\frac{0.005}{0.02}\right)}{0.005-0.002}=\frac{\ln(0.25)}{-0.015}=92.42\text{min}$$

于是，t_{max}时刻房室 I 中物质的最大含量为

$$q_1(t_{max})=\frac{q_2(0)K_3}{K_1+K_2-K_3}\left(e^{-K_3t}-e^{-(K_1+K_2)t}\right)\Big|_{t=92.42}$$

$$=\frac{50\times0.02}{0.005-0.02}(e^{-0.02\times92.42}-e^{-0.005\times92.42})=31.5\text{g}$$

■

12.2.3　多房室模型

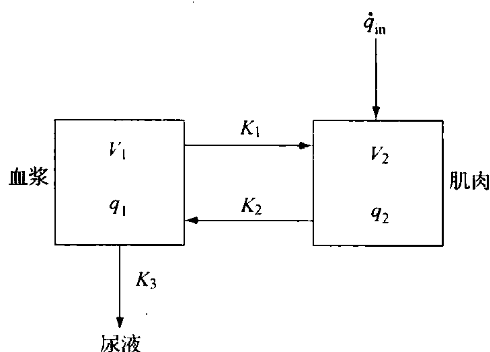
人体真实模型所包含的房室数目比上节所述模型要多，包括细胞房室、组织间隙房室和血浆房室等，并且这些房室单元还可以进一步分解。例如，组织间隙房室可以分成胃肠道、口腔、肝脏、肾脏等房室。其中每个房室都有其自己的物质传输率。通常，人们关注的并不是物质从一个房室传输到另一个房室的方式，而是物质传输的量。

上一节所述的建模方法适用于包含任意数目房室的模型。每个房室用一个质量守恒微分方程描述其物质变化速率。于是，包含 N 个房室的模型就有 N 个方程，其通式为

$$\frac{dq_i}{dt}=\text{输入}-\text{输出}$$

其中 q_i 为房室 i 中物质的量。线性系统的传输率为常数。

例 12.3　下图所示是一个描述人体内肌酐分布的双房室模型。肌酐是肌肉组织新陈代谢所产生的废物，通过尿液清除。设肌肉产生肌酐的速率为 $\dot{q}_{in}$，是模型的阶跃输入。请计算血浆房室中肌酐的浓度变化。



解：根据质量守恒定律，房室Ⅰ和房室Ⅱ中描述肌酐变化速率的微分方程分别为

$$\frac{dq_1}{dt}=K_2q_2-(K_1+K_3)q_1$$

$$\frac{dq_2}{dt}=K_1q_1-K_2q_2+\dot{q}_{in}$$

使用以下关系式将质量转化为浓度

$$浓度=\frac{质量}{体积}$$

可得

$$V_1\frac{dc_1}{dt}=K_2V_2c_2-(K_1+K_3)V_1c_1$$

$$V_2\frac{dc_2}{dt}=K_1V_1c_1-K_2V_2c_2+\dot{q}_{in}$$

下面用D算子消去变量 c_2。一般，Dy 表示一阶微分算子 $\frac{dy}{dt}$，D^2y 则表示 $\frac{d^2y}{dt^2}$，依此类推。利用D算子，可以通过代数方法消去变量，将微分方程组变成单个微分方程。这样，就可以求解微分方程。将以上微分方程组转化为D算子的形式，即为

$$V_1Dc_1=V_2K_2c_2-V_1(K_1+K_3)c_1$$

$$V_2Dc_2=V_1K_1c_1-V_2K_2c_2+\dot{q}_{in}$$

合并同类项，得

$$(V_1D+V_1(K_1+K_3))c_1-V_2K_2c_2=0$$

$$-V_1K_1c_1+(V_2D+V_2K_2)c_2=\dot{q}_{in}$$

为了消去 c_2，将第一个方程两边同乘以 $(V_2D+V_2K_2)$，并将第二个方程两边同乘以 V_2K_2，得

$$(V_2D+V_2K_2)(V_1D+V_1(K_1+K_3))c_1-(V_2D+V_2K_2)V_2K_2c_2=0$$

$$-V_2K_2V_1K_1c_1+V_2K_2(V_2D+V_2K_2)c_2=V_2K_2\dot{q}_{in}$$

将两个方程相加，消去 c_2，可得

$$(V_2V_1D^2+(V_2V_1K_2D+V_2V_1(K_1+K_3)D)+V_2V_1K_2K_3)c_1=K_2V_2\dot{q}_{in}$$

即

$$(D^2+(K_1+K_2+K_3)D+K_2K_3)c_1=\frac{K_2\dot{q}_{in}}{V_1}$$

将D算子重新用微分表示

$$\ddot{c}_1+(K_1+K_2+K_3)\dot{c}_1+K_2K_3c_1=\frac{K_2\dot{q}_{in}}{V_1}$$

此微分方程的特征根为

$$s_{1,2}=\frac{-(K_1+K_2+K_3)\pm\sqrt{(K_1+K_2+K_3)^2-4K_2K_3}}{2}$$

由于 K_1、K_2 和 K_3 都大于零，方程的自由响应是一个过阻尼响应，即

$$c_{1_n}=B_1e^{s_1t}+B_2e^{s_2t}$$

输入为常数，因此，强迫响应就是常数，设为 B_3，将其代入微分方程得到

$$c_{1_f}=B_3=\frac{\dot{q}_{in}}{K_3V_1}$$

于是，$t\geqslant0$ 时的全响应为

$$c_1=B_1e^{s_1t}+B_2e^{s_2t}+\frac{\dot{q}_{in}}{K_3V_1}$$

常数 B_1 和 B_2 可以利用初始条件求得。 ■

12.2.4 改进的房室模型

由上述内容可知，利用不同阶数的房室模型可以描述各种复杂的系统，并找到求解模型变量的方法。但是，如果系统的传输率不是常数，而是变量。例如，传输率是随着房室物质含量变化的变量，那么，上述房室分析法就不适用了。不过，可以用改进房室分析法的非线性模型来研究。非线性微分方程通常没有解析解，要用仿真方法求其数值解。还有一种处理非线性方程的方法，就是将非线性因素线性化，也就是加入伪稳态条件。

传染病模型

本节将通过传染病传播的问题，来介绍改进房室模型分析方法。流行病学是研究人群中疾病的产生及其各种影响因素的一门医学学科。建立传染病模型有助于

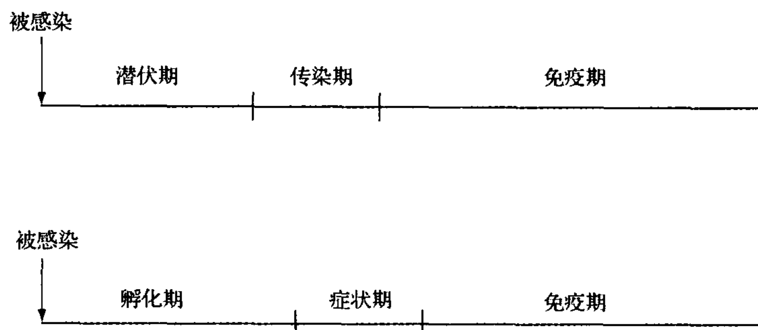
1）确定病因和病源；

2）控制疾病的传播。

历史上传染病模型的建立可以追溯到1760年，当时D. Bernoulli用数学模型研究了天花发病人口的动态变化过程。不过，20世纪以前，这方面的研究很少。到了20世纪初，Hamer和Soper提出了一种数学模型，用于描述苏格兰格拉斯哥（Glasgow，Scotland）所发生的麻疹疫情。1928年，Kermack和McKendrick用连续时间法，Reed和Frost用离散时间法，分别推进了Hammer和Soper的研究结果。20世纪50年代，Abbey和Bailey发表了相关研究工作，随后，这个领域的研究进展也好像传染病一样，一发不可收拾。之所以会形成这样的发展历史，主要是由于传染病

模型是非线性模型，必须依靠先进的计算工具和算法，才能求解其微分方程。

本节将介绍 Kermack-McKendrick 和 Reed-Frost 这两种传染病模型，每种模型所描述的疾病传染的时间过程有下面所示两种方式。



其中

1）潜伏期（Latent period）：从接触病菌到实际发病之间的时间间隔；

2）传染期（Infectious period）：病菌可以传染给其他人的时间段；

3）免疫期（Immune period）：人体暂时或者永久地对病菌具有免疫能力，不再传染病菌的时期；

4）孵化期（Incubation period）：从接触病菌到出现症状之间的时间间隔；

5）症状期（Symptomatic period）：病人具有明显症状的时间段。

Kermack-McKendrick 连续时间模型

设某地区总人口为 n，其中尚未感染的易感染者（susceptible）有 x 人，已感染者（infective）有 y 人，获得免疫能力（immune）的有 z 人，则

$$n=x+y+z$$

该模型具有如下假设：

1）各类人在总人口中均匀分布；

2）潜伏期为 0；

3）整个地区人群独立封闭，与外界隔离；

4）传染期呈负指数分布；

5）传染速率为 β；

6）清除速率为 γ（包括康复和死亡）。

独立封闭人群的急性传染过程与易感染者人数有关，也和易感染者与已感染者之间的传染速率有关。图 12.4 所示是房室模型框图，箭头表示各个状态之间的不可逆转换关系，转换速率由已感染者的传染速率 β 和清除速率 γ 决定。

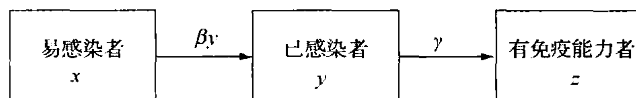


图 12.4 Kermack-McKendrick 房室模型

Kermack-McKendrick 模型描述了 t 时刻易感染者 x、已感染者 y 和有免疫能力者 z 之间转换的关系，其传染速率为 βy，该模型的微分方程组为：

$$\frac{dx}{dt}=-\beta xy$$

$$\frac{dy}{dt}=\beta xy-\gamma y$$

$$\frac{dz}{dt}=\gamma y$$

幸运的是，利用泰勒级数近似法就可以求解此非线性方程组的解析解。假设 $t=0$ 时刻，$y_0=y(0)$ 和 $z_0=z(0)$ 的值都很小，也就是 $x_0=x(0)\approx n$。定义相对清除速率为

$$\rho=\frac{\gamma}{\beta}$$

为了使$\frac{dy}{dt}>0$，必须有 $\rho<x_0$；否则，如果 $\rho>x_0$，传染病就不会流行。因此，相对清除速率 $\rho=x_0$ 所对应的易受感染者人数就是临界阈值。

将第一个方程除以第三个方程，消去 y，可得微分方程

$$\frac{\frac{dx}{dt}}{\frac{dz}{dt}}=\frac{-\beta xy}{\gamma y}=\frac{-x}{\rho}$$

求解该方程，可得 x 为

$$x=x_0 e^{\frac{-z}{\rho}}$$

将此 x 解和 $y=n-x-z$ 代入第三个微分方程，可以求 z 的解，即

$$\frac{dz}{dt}=\gamma y=\gamma(n-x-z)=\gamma\left(n-z-x_0 e^{\frac{-z}{\rho}}\right)$$

利用泰勒级数展开，此式变成

$$\frac{dz}{dt}=\gamma\left\{n-z-\underbrace{x_0+\frac{x_0 z}{\rho}-\frac{x_0}{2}\frac{z^2}{\rho^2}+\cdots}_{\text{泰勒级数}}\right\}$$

舍去高阶项，可得

$$\frac{dz}{dt}=\gamma\left\{(n-x_0)+\left(\frac{x_0}{\rho}-1\right)z-\frac{x_0}{2\rho^2}z^2\right\}$$

利用 $\tanh(x)=\frac{e^x-e^{-x}}{e^x+e^{-x}}$进行变换之后，可得 z 的解为

$$z=\frac{\rho^2}{x_0}\left\{\frac{x_0}{\rho}-1+\alpha\tanh\left(\frac{1}{2}\alpha\gamma t-\phi\right)\right\} \tag{12.21}$$

其中

$$\alpha=\sqrt{\frac{(x_0-1)^2}{\rho}+\frac{2x_0 y_0}{\rho^2}}$$

且

$$\phi=\operatorname{arctanh}\frac{1}{\alpha}\left(\frac{x_0}{\rho}-1\right)$$

为了计算传染病最终所波及的总人数，下面用式（12.21）求 $z(+\infty)$。由于 $\tanh\left(\frac{1}{2}\alpha\gamma t-\phi\right)_{t\to\infty}=1$，得

$$z(+\infty)=\frac{\rho^2}{x_0}\left\{\frac{x_0}{\rho}-1+\alpha\right\}$$

注意，如果

$$\left(\frac{x_0}{\rho}-1\right)^2\gg\frac{2x_0y_0}{\beta},$$

那么

$$\alpha=\left(\frac{x_0}{\rho}-1\right)$$

且

$$z(+\infty)\approx2\rho\left(1-\frac{\rho}{x_0}\right)$$

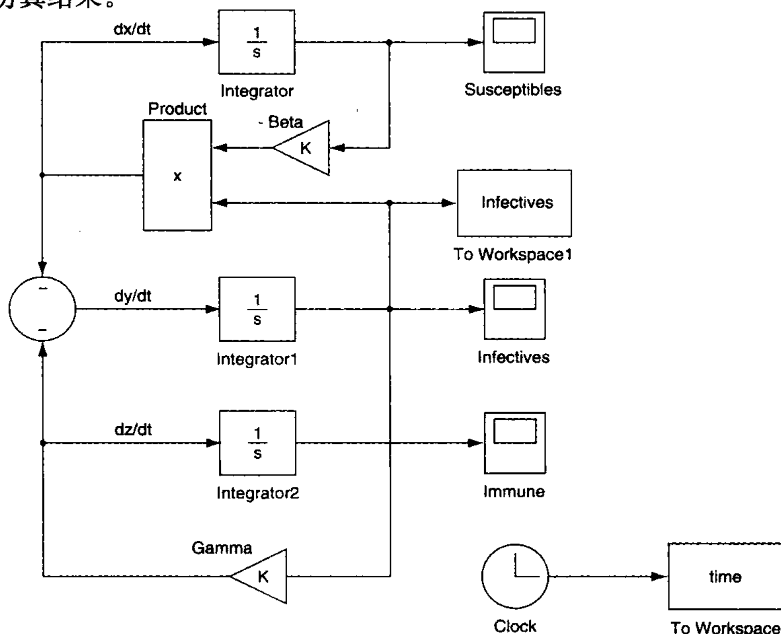
当 $x_0<\rho$ 时，传染病根本不会流行。

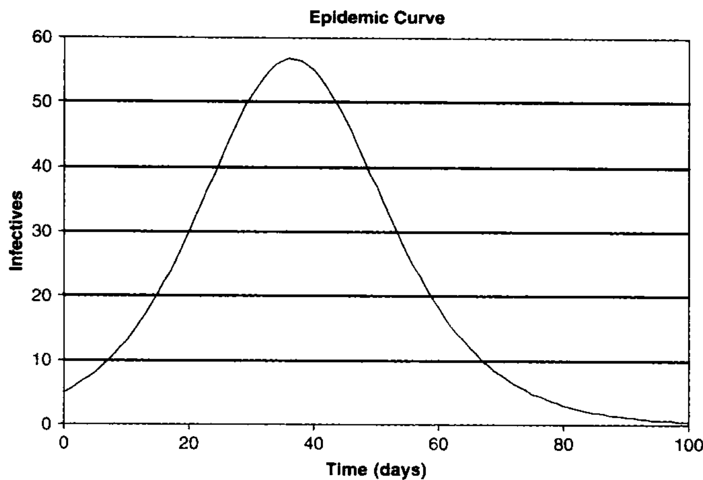
例 12.4　假设某地区人口中易感染者有 10000 人，已感染者有 5 人，没有已获免疫者，且 $\beta=0.0001$，$\gamma=0.9$。请利用 Kermack-McKendrick 模型建立一个 SIMULINK 仿真程序，并作出已感染者人数变化曲线，即传染病曲线（epidemic curve）。

解：建立仿真模型之前，首先考察该传染病是否会流行，计算相对清除速率 ρ

$$\rho=\frac{\gamma}{\beta}=\frac{0.90}{0.0001}=9000\leqslant10000$$

可见 ρ 小于初始易感染者人数，因此，传染病将流行起来。下图是 SIMULINK 程序模型及其仿真结果。





■

如果要长期跟踪患病人口的变化情况，那么，如图 12.5 所示，模型中还必须加入人口出生率和死亡率。于是，该模型的微分方程组就变成

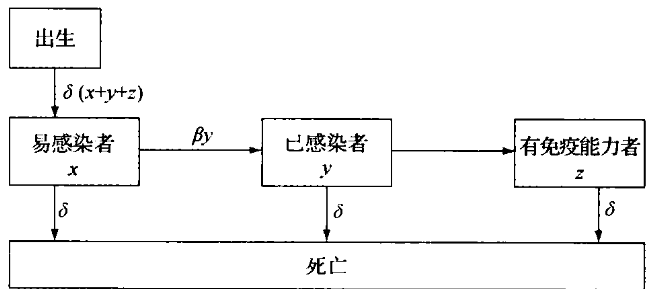


图 12.5 考虑出生率和死亡率的 Kermack-McKendrick 房室模型

$$\frac{dx}{dt}=-\beta xy+\delta(z+y)$$

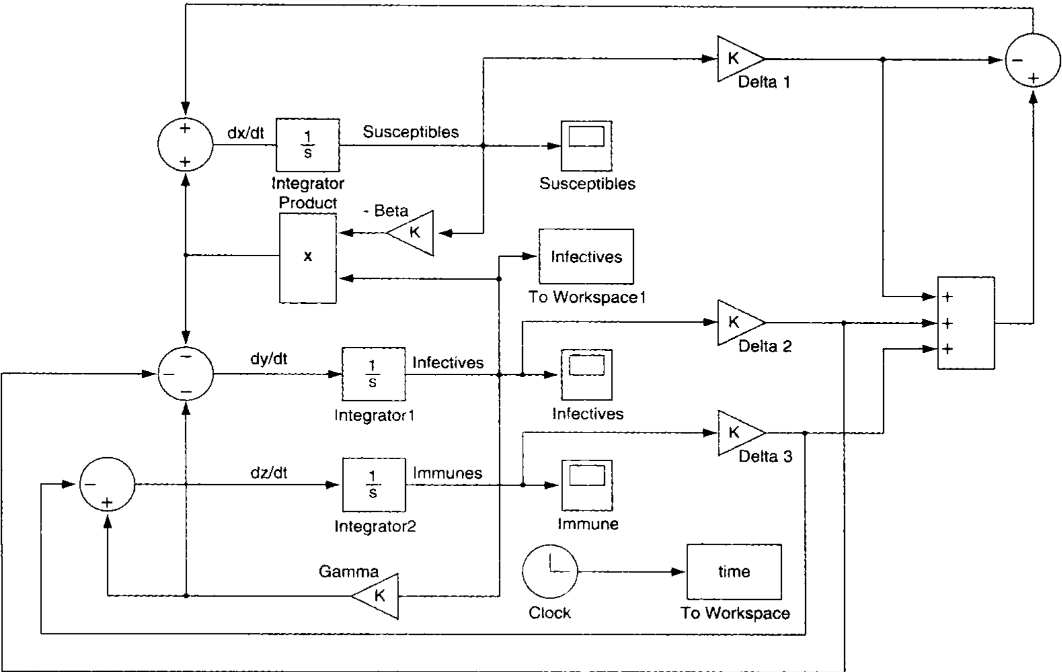
$$\frac{dy}{dt}=\beta yx-\gamma y-\delta y$$

$$\frac{dz}{dt}=\gamma y-\delta z$$

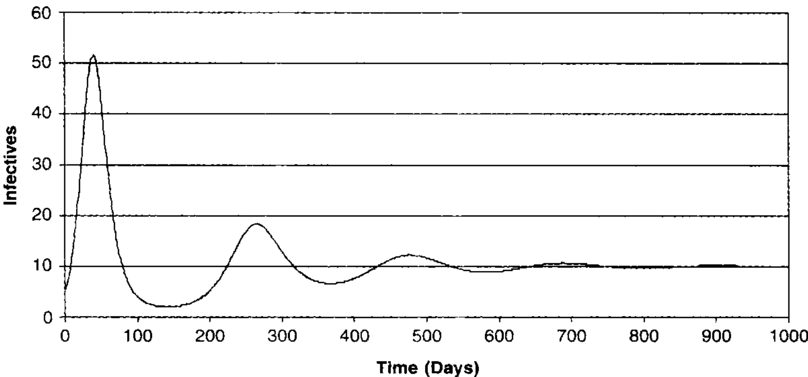
但是，不幸的是，模型中包括死亡率之后，传染病就变成了地方性疾病，也就是该地区始终保持一定的已感染者人数，疾病不会消失。下面举例说明这种情况。

例 12.5 假设某地区人口中易感染者有 10000 人，已感染者有 5 人，没有已获免疫者，且 $\beta=0.0001$，$\gamma=0.9$，$\delta=0.01$。请应用包括出生和死亡的 Kermack-McKendrick 模型建立一个 SIMULINK 仿真模型，并作出传染病曲线。

解： SIMULINK 模型如下



下图的仿真结果显示了该疾病的地方病特征。但是，必须注意，实际这种疾病的发展过程并非如此。



■

Reed-Frost 模型

Reed-Frost 模型是一种确定性的离散时间模型。不用连续时间模型，而用离散时间模型，其原因之一是测量数据总是按照一定的时间间隔采集。还有一个原因是离散时间模型容易实现简单模型的扩展，例如增加具有给定分布的非零潜伏期等。此

模型的假设有

1）各类人在总人口中均匀分布；

2）潜伏期不为 0；

3）整个地区人群独立封闭，并处于稳态；

4）所有易感染者接触已感染者之后都会被传染，并只在紧接的一段时间内会向其他人传播，此后即获得免疫能力。

5）以上所述这段时间就是传染期，人在传染期内任何时刻都可能被传染，因此平均潜伏期为整个传染期的一半。

6）在一段时期内，每个人与其他任何特定个体的有效接触概率都一样，设为常数 p。

Reed-Frost 模型的结构如图 12.6 所示。注意，有效接触概率 p 可以定义为

$$p=\frac{\text{有效接触平均数}}{N}$$

设

$$q=1-p$$

那么，易感染者没有得到有效接触的概率为

$$q^{C_k}$$

Reed-Frost 模型描述了离散时间 $k+1$ 下，易感染者 S、已感染者 C 和免疫者 R 之间的状态转换关系。在某个给定时期内，易感染者与已感染者发生有效接触之后，将会被传染，并只在紧接的一段时期会再传染给其他人，然后，就成为免疫者。由于感染速率为（$1-q^{C_k}$），此模型的非线性差分方程组为

$$C_{k+1}=S_k\ (1-q^{C_k})$$
$$S_{k+1}=S_k-C_{k+1}$$
$$R_{k+1}=R_k+C_k$$

离散时间间隔（即采样周期 T）设定为与传染期一样长，因此，清除率就等于 1。

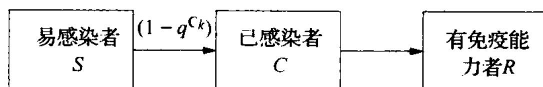
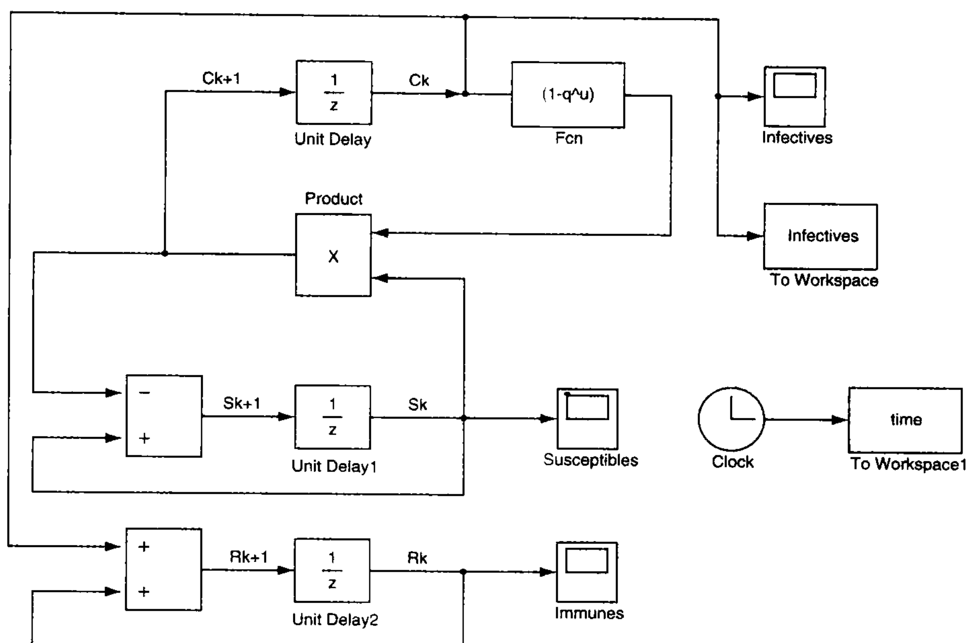


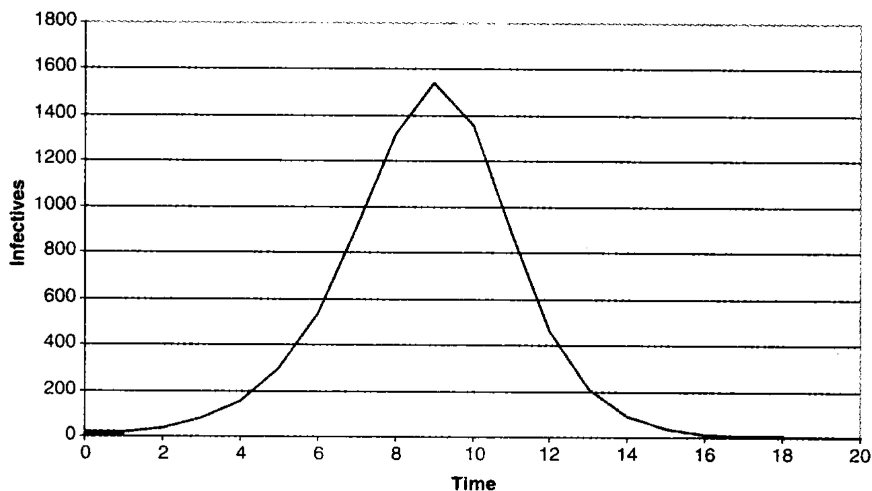
图 12.6　Reed-Frost 模型

例 12.6　假设某地区人口中易感染者有 10000 人，已感染者有 10 人，没有已获免疫者，且 $q=0.9998$。请利用 Reed-Frost 模型建立一个 SIMULINK 仿真模型，并作出传染病曲线。

解： SIMULINK 模型如下



仿真结果如下



■

12.2.5 各种生理房室之间流体物质的传输

本节我们用改进房室模型来分析房室之间流体物质的传输。图 12.7 所示是液体流过某个房室的示意图，其中，V_A为房室的容积，$\frac{dq_{in}}{dt}=\dot{q}_{in}$为物质进入房室的速率，$\frac{dq_{out}}{dt}=\dot{q}_{out}$为物质流出房室的速率。

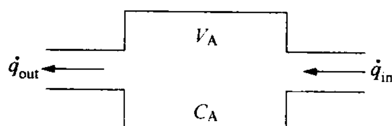


图 12.7 描述房室之间流体物质传输的房室模型

下面根据质量守恒定律分析这种系统。物质流入房室的速率为

$$\dot{q}_{in}=\dot{V}c_{in}$$

式中 $\dot{V}$——液体流入房室的容积速率；

c_{in}——流入房室的液体中物质的浓度。

物质流出房室的速率为

$$\dot{q}_{out}=\dot{V}c_{out}$$

式中 c_{out}——流出房室的液体中物质的浓度。

描述房室内物质质量变化速率的微分方程为：

$$\frac{dq_A}{dt}=\dot{q}_{in}-\dot{q}_{out}=\dot{V}(c_{in}-c_{out})$$

改用浓度形式表示，即为

$$V_A\frac{dc_A}{dt}=\dot{V}(c_{in}-c_{out})$$

这就是描述房室之间流体物质传输的基本方程。如果所描述的是心血管系统，那么，就用心输出量 $\dot{Q}$ 代替容积速率 $\dot{V}$。

例 12.7 下图所示是描述动脉血和静脉血中类固醇浓度随时间变化的三房室模型。类固醇以团注形式被快速注入肺动脉。假设所有类固醇的新陈代谢都在肝脏中进行，并且其代谢速率直接与组织房室中类固醇的浓度成正比。$\dot{Q}$ 为心输出量。如果心输出总量的 5%流入肝动脉，20%流入门静脉，请写出描述此系统的微分方程。其中

$A\delta(t)$——类固醇团注；

$\dot{Q}$——心输出量；

c_p——肺中类固醇的浓度；

c_L——肝脏中类固醇的浓度；

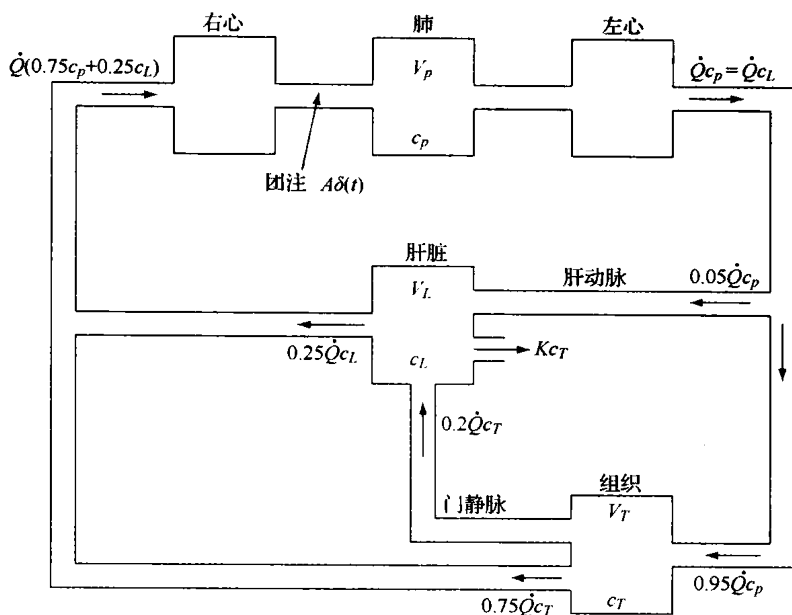
c_T——组织中类固醇的浓度；

V_p——肺房室容积；

V_L——肝房室容积；

V_T——组织房室容积；

K——代谢率系数。



解：每个房室的质量守恒方程如下：

肺房室：

$$V_p\frac{\mathrm{d}c_p}{\mathrm{d}t}=\dot{Q}(0.75c_T+0.25c_L)+A\delta(t)-\dot{Q}c_p$$

肝脏房室：

$$V_L\frac{\mathrm{d}c_L}{\mathrm{d}t}=0.05\dot{Q}c_p+0.2\dot{Q}c_T-0.25\dot{Q}c_L-Kc_T$$

组织房室：

$$V_T\frac{\mathrm{d}c_T}{\mathrm{d}t}=0.95\dot{Q}c_p-0.95\dot{Q}c_T$$ ■

12.2.6　染料稀释模型

染料稀释法可以用于测定心输出量、心脏功能、各种器官的灌注功能以及血管系统的状况等。一般，将 Evans 蓝等染料从心血管系统的某个部位注入，然后，在一个或多个位点检测染料浓度随时间的变化过程。

显然，染料必须能够在血液中停留较长时间。如果测量过程中有部分染料离开血浆，那么，测量结果就会出现误差。最常用的染料是 Evans 蓝，它能够与血浆中的白蛋白结合。还有一种是放射性铬同位素 Cr^{51}。

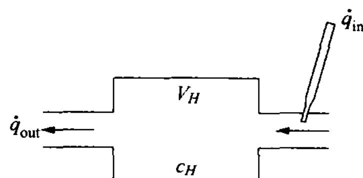
例 12.8　请计算人体的血容量。方法是：将已知质量的染料注入血流中，几分钟之后，等到染料在血液中混合均匀，再抽取血液样本，经检测后用于计算。

解：

$$\text{血容量}=\frac{\text{注入的染料质量}}{\text{血液的染料浓度}}$$ ■

例 12.9 请计算心输出量。

解：如下图所示，建立系统模型时，将心脏和肺脏看作一个流通房室，染料注射设定为幅值为 K 的阶跃输入。



图中 V_H——心肺房室的容积；

$\dot{q}_{in}$——染料注射速率；

$\dot{q}_{out}$——染料离开房室的速率；

c_H——心肺房室中的染料浓度。

根据质量守恒定律，有

$$\frac{\mathrm{d}q_H}{\mathrm{d}t}=\dot{q}_{in}-\dot{q}_{out}$$

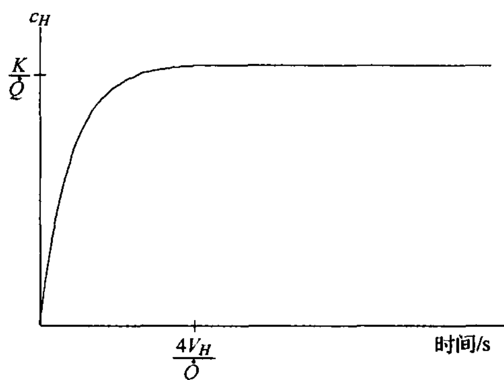
将 $\dot{q}_{out}=\dot{Q}c_H$，以及阶跃染料输入 $\dot{q}_{in}=Ku(t)$ 代入此式，可得

$$V_H\frac{\mathrm{d}c_H}{\mathrm{d}t}=\dot{q}_{in}-\dot{Q}c_H=Ku(t)-\dot{Q}c_H$$

该微分方程在 $t\geqslant 0$ 时的解为

$$c_H(t)=\frac{K}{\dot{Q}}\left(1-\mathrm{e}^{-\frac{\dot{Q}t}{V_H}}\right)$$

其曲线如下图所示。



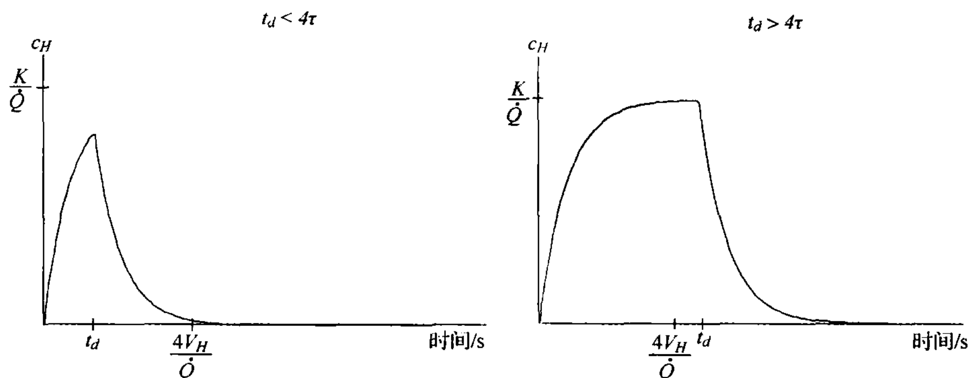
由图中曲线可知，根据实验数据，可以先从稳态值中求得心输出量 $\dot{Q}$，然后，从时间常数求得心肺房室的容积 V_H。然而，不幸的是，不经过心血管系统的多轮循环，该系统不会达到稳态，因此，这种方法并不能正确计算出心输出量。

如果染料注射不是阶跃式的，而是脉冲式的团注，那么，染料浓度的解变为

$$c_H(t)=\frac{K}{\dot{Q}}\left(1-e^{-\frac{\dot{Q}t}{V_H}}\right)u(t)-\frac{K}{\dot{Q}}\left(1-e^{-\frac{\dot{Q}(t-t_d)}{V_H}}\right)u(t-t_d)$$

式中　t_d——脉冲宽度。

根据脉冲宽度的不同，这个解可以呈现下图所示的两种形式。



由于心血管系统在不断循环，很难从实际测量数据中计算出心输出量 $\dot{Q}$。为了解决这个问题，设定染料注射的脉冲宽度 $t_d<4\tau$（τ 为时间常数），也就是不受再循环影响，那么，通过计算曲线下所包围的面积，可以求浓度的积分。即

$$\begin{aligned}S&=\int_0^\infty c_H(t)\mathrm{d}t\\&=\int_0^\infty\frac{K}{\dot{Q}}\left(1-e^{\frac{-\dot{Q}t}{V_H}}\right)\mathrm{d}t-\int_{t_d}^\infty\frac{K}{\dot{Q}}\left(1-e^{\frac{-\dot{Q}(t-t_d)}{V_H}}\right)\mathrm{d}t\\&=\frac{K}{\dot{Q}}\left(\int_0^{t_d}\mathrm{d}t-\int_0^\infty e^{\frac{-\dot{Q}t}{V_H}}\mathrm{d}t+\int_{t_d}^\infty e^{\frac{-\dot{Q}(t-t_d)}{V_H}}\mathrm{d}t\right)\\&=\frac{K}{\dot{Q}}\left[t_d+\frac{V_H}{\dot{Q}}-\frac{V_H}{\dot{Q}}\right]=\frac{K}{\dot{Q}}t_d\end{aligned}$$

这样，就可以求得 $\dot{Q}$ 为

$$\dot{Q}=\frac{Kt_d}{S}$$

■

12.3　快速眼动系统概述

快速眼动常称为扫视（Saccade），是指视线快速从视野内一个注视点跳到另一个注视点时眼球的运动。扫视这个词来自法语，原意是急速勒马。这种眼球运动发生的频率很高，阅读时尤其多，当眼睛看到一行末尾时，就会快速移到下一行开始处。

眼睛的眼动系统控制各种刺激引起的所有眼球运动，扫视只是其中的一种。眼动系统控制眼球在眼眶内转动，从而使人可以定位、注视、或者跟踪视野中的物体。眼球可以在眼眶内产生垂直、水平和旋转3个方向的转动，这些转动分别由3对主动肌一拮抗肌完成。主动肌和拮抗肌的作用相反，并且遵循交互神经支配原理。图12.8显示了控制眼球活动的几对肌肉。该控制系统的最终目标是将视网膜上的中央凹对准视野中的感兴趣目标。

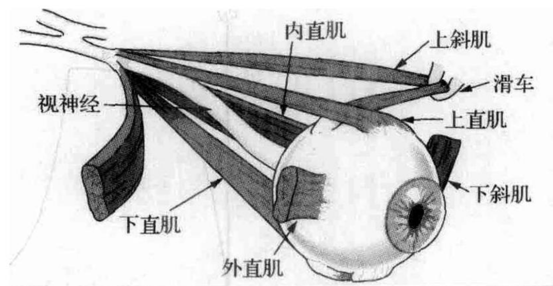


图12.8 右眼的肌肉和视神经

注：左眼与其类似，只是内直肌和外直肌的位置相反。其中，内直肌和外直肌一起控制眼球的水平方向转动；上直肌和下直肌控制眼球的垂直方向运动；上斜肌和下斜肌则控制眼球的旋转。每块肌肉作用力的大小由眼球所处的位置决定。眼睛平直注视正前方的位置称为第一眼位，此时各块肌肉在神经刺激下处于紧张状态。

眼动系统接受并响应来自视觉、听觉和前庭的刺激信号，可以产生以下5种眼球运动：用于捕捉目标的快速扫视运动；用于跟踪目标的平稳追踪运动；头部转动时，用于将视线保持在目标上的前庭眼动反射；用于注视较近或较远目标时的会聚或辐散运动；最后还有视动运动，当视线扫过充满目标的环境时，或者在头部持续转动过程中，仍然要保持目标影像落在中央凹上，这时眼球就会产生视动运动。除了会聚一辐散运动之外，其他4种眼球运动都是共轭的，也就是，两个眼球一起运动，方向和距离都一样。会聚一辐散眼动是非共轭眼球运动，用于将视线保持在远近不同距离的目标上。如果目标移近，则两眼视轴就会聚起来；反之，如果目标远离，则两眼视轴会发散。以上这些眼球运动受到不同的神经系统控制，但最终都通过相同的途径作用于眼部肌肉。诱发这5种眼动的刺激还会引起头部和身体的其他运动。因此，视觉系统属于多输入多输出系统的一部分。

无论输入情况如何，眼动系统的主要任务就是通过眼球的运动使被注视目标聚焦到视网膜中间$\frac{1^\circ}{2}$角度区域，此区域称为中央凹（见图12.9）。视网膜内排列的感光细胞可以将呈现在视网膜上的图像转化为神经脉冲信号，这些脉冲信号再沿着视神经传向中枢神经系统，经过神经元平行通路，到达上丘脑和大脑皮层。与其周边视网膜相比，中央凹分布的感光细胞密度较高，因此，中央凹区域具有较高的分辨率，即较高的视敏度。中央凹的作用是看清楚目标物体，而视网膜周边区域的作用则是

探测新目标。一旦周边区域探测到了新目标，眼动系统就会尽快将眼球重新定位到该目标上。这种扫视通常称为目标定向扫视。

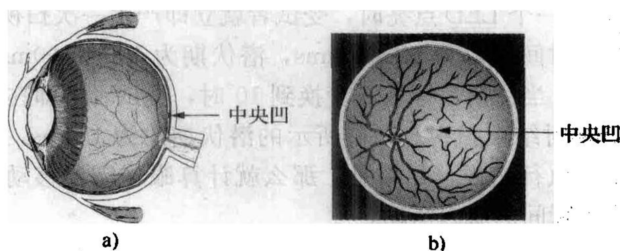


图 12.9 眼球的剖面图（图 a）眼球后面是视网膜，视网膜属于中枢神经系统的一部分，由视杆细胞和视锥细胞两种感光细胞组成；从眼睛前方看进去的视网膜眼底图（图 b），中央凹居中，直径约 1mm，眼动系统就是要将所注视的目标对准中央凹

扫视过程中，眼动系统工作于开环模式。扫视过后，系统再工作于闭环模式，将眼球调整到正确的位置。因此，扫视时的快速眼动并不存在反馈，这是因为来自视网膜和肌肉感受器的信号传导速度不够快，不可能在快速扫视期间用于调节眼动。眼睛的扫视由眼动系统和扫视控制器两个基本组成部分共同完成。眼动系统包含 3 对眼肌和眼球。这 3 对眼肌的伸缩可以控制眼睛在水平、垂直和旋转 3 个方向上的转动（见图 12.8）。每对眼肌都在扫视控制器的交互神经支配下，以相互拮抗的方式产生作用。为了简化，本书只介绍眼球水平转动的模型，它涉及 1 对眼肌，即外直肌和内直肌。

12.3.1 扫视眼动的特征

眼睛的扫视是人类能够产生的最快的随意肌运动，它可以将眼睛的注视点从一个位置快速移动到另一个位置。图 12.10 所示是一个幅度为 10°的扫视记录信号。扫

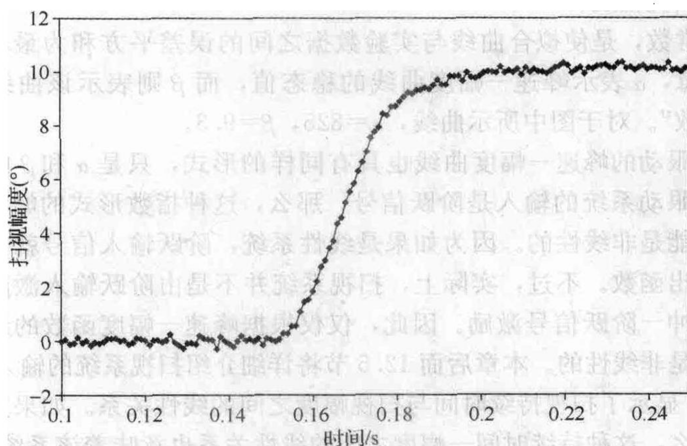


图 12.10 幅度约为 10°的一次扫视眼动记录（数据采样率为 1000 Hz）

视实验记录过程中，通常让受试者坐在水平目标显示器前，显示器由小型发光二极管（即LED）组成。受试者根据要求，尽可能快速移动眼睛，准确盯住闪光的LED。当某个LED熄灭，另一个LED点亮时，受试者就立即产生一次扫视。扫视眼动具有共轭弹道性质，持续时间通常为30～100ms，潜伏期为100～300ms。如图12.10所示，受试者正视前方，当注视目标从0°变换到10°时，受试者随即在150ms后启动一次扫视，并于200ms时结束扫视。图中所示的潜伏期约为150ms，这段时间内，中枢神经系统决定是否执行扫视，如果是，那么就计算眼球需要移动的距离，并将视网膜上的视差转化为瞬间的肌肉活动。

一般，眼睛每次执行的扫视差别很大，其潜伏期、扫视峰速（即扫视的最大速度）、到达峰速的时间、扫视持续时间等都可以在很大范围内变动。而且，对于同样的扫视幅度，各个变量的变化能够很好地相互协调。例如，峰速较小的扫视具有较长的持续时间；而峰速较高的扫视则具有较短的持续时间。这样，到达同一位置的各次扫视往往具有不同的运动轨迹。

利用扫视的几个主特性曲线可以分析扫视的动态过程，这些特性曲线包括峰速—幅度曲线、持续时间—幅度曲线和潜伏期—幅度曲线。图12.11所示是某受试者26次不同扫视幅度的主特性曲线。该受试者实际上在正负两个方向上共执行了52次扫视，为了简化，图中只显示了正向扫视结果。注意，峰速－幅度曲线的起始部分几乎是直线上升，而后，对于幅度较大的扫视，峰速趋于平稳，几乎变成常数。很多人用指数函数拟合这种关系，图12.11a中的实线就是正向扫视眼动峰速－幅度数据的指数拟合，其拟合函数的方程为

$$V=\alpha\left(1-e^{-\frac{x}{\beta}}\right) \tag{12.22}$$

式中 V——扫视峰速；

x——扫视幅度；

α和β——常数，是使拟合曲线与实验数据之间的误差平方和为最小的数值。注意，α表示峰速—幅度曲线的稳态值，而β则表示该曲线的“时间常数”。对于图中所示曲线，$\alpha=825$，$\beta=9.3$。

负向扫视眼动的峰速—幅度曲线也具有同样的形式，只是α和β的数值有所不同。如果扫视眼动系统的输入是阶跃信号，那么，这种指数形式的峰速—幅度曲线表明该系统可能是非线性的。因为如果是线性系统，阶跃输入信号就会产生线性的峰速—幅度输出函数。不过，实际上，扫视系统并不是由阶跃输入激励的，而是由比较复杂的脉冲－阶跃信号激励。因此，仅仅根据峰速—幅度函数的形式并不能判定扫视系统就是非线性的。本章后面12.5节将详细介绍扫视系统的输入激励信号。

图12.11b显示了扫视持续时间与扫视幅度之间的线性关系。如果系统的输入为阶跃信号，那么，这种持续时间—幅度之间的线性关系也意味着该系统是非线性的。因为线性系统的阶跃输入响应总是具有恒定不变的持续时间。不过，由于扫视系统的输入并不是阶跃函数，因此，也不能仅仅根据持续时间—幅度曲线的形式来判定

扫视系统是否是非线性的。

图 12.11c 所示是根据潜伏期和扫视幅度两个参数作出的数据点。显然，潜伏期与扫视幅度之间没有什么关系，也就是潜伏期的大小与扫视幅度无关。但是，有些研究人员却认为潜伏期与扫视幅度之间存在线性关系。本章下面所建立的眼动模型中，潜伏期是隐含参数，不是主要参数，因此，不重要。

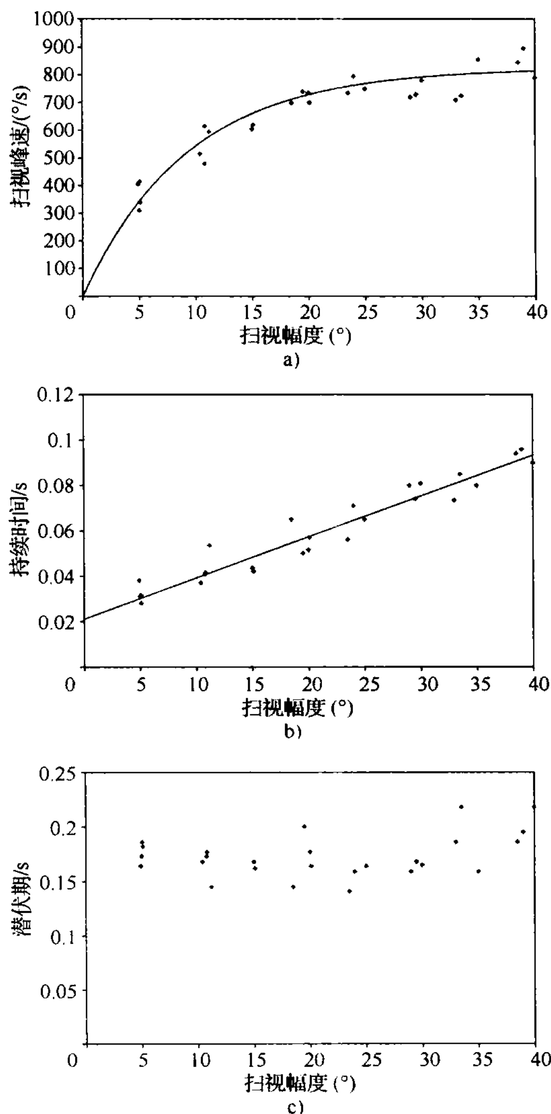


图 12.11　正向扫视眼动的主特性曲线（摘自 Enderle，Observations on pilot neurosensory control performance during saccadic eye movements. Aviation，Space，and Environmental Medicine，59：309，1988.）

注：某受试者不同幅度的 26 次扫视的峰速一扫视幅度曲线（图 a），持续时间一扫视幅度曲线（图 b），潜伏期一扫视幅度曲线（图 c）。负向扫视的特性曲线与此类似。

由于眼动系统很复杂，这里我们只研究水平方向的快速眼动。实事上，眼球有水平、垂直和旋转 3 个方向的运动。眼动系统模型包括与眼球相连的每块肌肉的模块，以及控制每对肌肉的独立扫视控制器模块。本章所述的几种水平眼动模型都已有一定的研究历史，这里将按照复杂度逐渐增加的顺序逐一介绍不同的模型，这样，便于读者全面了解各种模型的重要意义。但并不是介绍所有眼动模型，只是列举少数几个例子。

12.4 Westheimer 扫视眼动模型

图 12.12 所示是 1954 年 Westheimer 发表的第一个水平眼动扫视定量模型。通过考察 20°扫视的实验记录数据，并假定扫视控制器为阶跃输入信号，Westheimer 用如下二阶微分方程描述了图中所示模型：

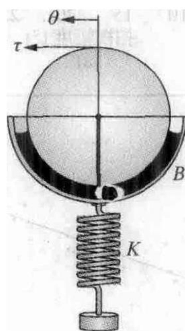


图 12.12 Westheimer 二阶扫视系统模型示意图

注：参数 J、B 和 K 分别为转动惯量、摩擦系数和刚性系数，用于表示眼球及其附件的粘弹性。内直肌和外直肌施加在眼球上的力矩用 $\tau(t)$ 表示，θ 表示眼球所处位置的角度。眼球半径设为 r。

$$J\ddot{\theta}+B\dot{\theta}+K\theta=\tau(t) \tag{12.23}$$

下面用拉普拉斯变换分析该模型的特性，并与实验数据进行比较。

对上式求拉普拉斯变换，并设初始条件为 0，可得

$$s^2J\theta+sB\theta+K\theta=\tau(s)$$

$$(s^2J+sB+K)\theta=\tau(s) \tag{12.24}$$

其传递函数的标准形式为

$$H(s)=\frac{\theta(s)}{\tau(s)}=\frac{\frac{\omega_n^2}{K}}{s^2+2\zeta\omega_n s+\omega_n^2} \tag{12.25}$$

式中，$\omega_n=\sqrt{\frac{K}{J}}$，且 $\zeta=\frac{B}{2\sqrt{KJ}}$。根据 20°扫视的眼动轨迹，Westheimer 计算得到 $\omega_n=120\text{rad/s}$，且 $\zeta=0.7$。当阶跃输入为 $\tau(s)=\frac{\gamma}{s}$ 时，可求得 $\theta(t)$ 为：

$$\theta(t)=\frac{\gamma}{K}\left[1+\frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}}\cos(\omega_d t+\phi)\right] \tag{12.26}$$

式中，$\omega_d=\omega_n\sqrt{1-\zeta^2}$，$\phi=\pi+\arctan\dfrac{-\zeta}{\sqrt{1-\zeta^2}}$。

例 12.10　请写出从式（12.25）推导式（12.26）的中间步骤。

解：将阶跃输入 $\tau(s)=\dfrac{\gamma}{s}$ 代入式（12.25），得到

$$\theta(s)=\frac{\gamma\omega_n^2}{Ks(s^2+2\zeta\omega_n s+\omega_n^2)}$$

根据 Westheimer 的估计，此式分母含有一对复数根，因此，将其进行部分分式展开，可得

$$\theta(s)=\frac{\gamma}{Ks}+\frac{\dfrac{\gamma}{2K((\zeta^2-1)-j\zeta\sqrt{1-\zeta^2})}}{s+\zeta\omega_n-j\omega_n\sqrt{1-\zeta^2}}+\frac{\dfrac{\gamma}{2K((\zeta^2-1)+j\zeta\sqrt{1-\zeta^2})}}{s+\zeta\omega_n+j\omega_n\sqrt{1-\zeta^2}}$$

$$=\frac{\gamma}{Ks}+\frac{|M|e^{j\phi}}{s+\zeta\omega_n-j\omega_n\sqrt{1-\zeta^2}}+\frac{|M|e^{-j\phi}}{s+\zeta\omega_n+j\omega_n\sqrt{1-\zeta^2}}$$

$|M|$ 为部分分式系数（即两个复数项的分子）的模值，其大小为：

$$|M|=\frac{\gamma}{2K\sqrt{(\zeta^2-1)^2+\zeta^2(1-\zeta^2)}}=\frac{\gamma}{2K\sqrt{1-\zeta^2}}$$

ϕ 为相角，其求法如下：将 $\theta(s)$ 第二个部分分式系数的分子和分母同乘以分母的共轭复数，从而去掉分母的虚部，并重排，可得

$$\frac{\gamma}{2K((\zeta^2-1)-j\zeta\sqrt{1-\zeta^2})}=\frac{\gamma((\zeta^2-1)+j\zeta\sqrt{1-\zeta^2})}{2K((\zeta^2-1)-j\zeta\sqrt{1-\zeta^2})((\zeta^2-1)+j\zeta\sqrt{1-\zeta^2})}$$

$$=\frac{\gamma(-(\sqrt{1-\zeta^2})^2+j\zeta\sqrt{1-\zeta^2})}{2K((\zeta^2-1)-j\zeta\sqrt{1-\zeta^2})((\zeta^2-1)+j\zeta\sqrt{1-\zeta^2})}$$

$$=\gamma\sqrt{1-\zeta^2}\frac{(-\sqrt{1-\zeta^2}+j\zeta)}{2K((\zeta^2-1)-j\zeta\sqrt{1-\zeta^2})((\zeta^2-1)+j\zeta\sqrt{1-\zeta^2})}$$

如下图所示，由此式可求得部分分式系数的相角为

$$\phi=\pi+\psi=\pi+\arctan\frac{\zeta}{-\sqrt{1-\zeta^2}}$$

即

$$\psi=\arctan\left(\frac{\zeta}{-\sqrt{1-\zeta^2}}\right)$$

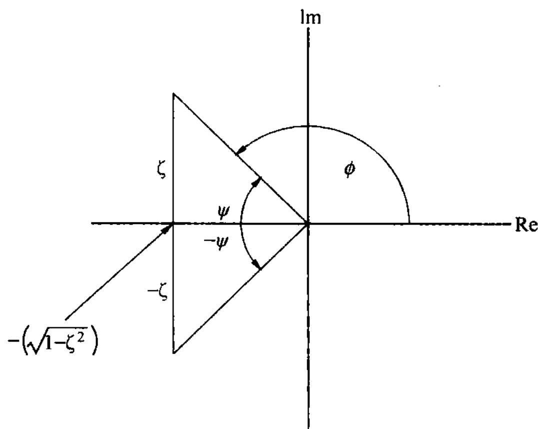
注意，图中直角三角形的斜边长度为 1。返回到时域即可得式（12.26）。注意，两个复数项的解为

$$2|M|e^{-\zeta\omega_n t}\cos\left(\omega_n\sqrt{1-\zeta^2}t+\phi\right)$$

通过求初始 0 时刻的 θ 值，可以简单验证式（12.26）**的正确性，即**

$$\theta(0)=\frac{\gamma}{K}\left[1+\frac{\cos(\phi)}{\sqrt{1-\zeta^2}}\right]=\frac{\gamma}{K}\left[1+\frac{-\sqrt{1-\zeta^2}}{\sqrt{1-\zeta^2}}\right]=0$$

扫视开始时，眼球处于平直注视正前方的第一眼位（primary position），也就是 $\theta(0)=0$。



■

为了全面考察模型的质量，必须将其仿真结果与实验数据进行比较。对于眼动系统，实验中比较容易测得的参数就是扫视达到最大角度所需的时间（即扫视持续时间）以及峰速。本章 12.3.1 节已经介绍过这两个参数。

下面计算扫视达到最大角度的时间 T_p。先应用求导的链式法则求式（12.26）的导数，即

$$\frac{\partial\theta}{\partial t}=\frac{\gamma e^{-\zeta\omega_n t}}{K\sqrt{1-\zeta^2}}[-\zeta\omega_n\cos(\omega_d t+\phi)-\omega_d\sin(\omega_d t+\phi)] \tag{12.27}$$

然后，设 $\left.\frac{\partial\theta}{\partial t}\right|_{t=T_p}=0$，计算 T_p，有

$$T_p=\frac{\pi}{\omega_n\sqrt{1-\zeta^2}} \tag{12.28}$$

根据 Westheimer 模型给定的参数值，对于任意幅度的扫视，都有 $T_p=37\text{ms}$，也就是，扫视持续时间 T_p 的值与扫视幅度的大小无关，这不符合图 12.11 所示的实验数据。实验数据表明扫视持续时间随着扫视幅度的增大而增加。

例 12.11 请由式（12.27）等于 0 推导出式（12.28），并求最大扫视角度 $\theta(T_p)$。

解： 将式（12.27）置 0，有

$$\frac{\gamma e^{-\zeta\omega_n t}}{K\sqrt{1-\zeta^2}}[-\zeta\omega_n\cos(\omega_d t+\phi)-\omega_d\sin(\omega_d t+\phi)]=0$$

由于方括号外面的乘积项不会等于 0，将其去除，可得

$$-\zeta\omega_n\cos(\omega_d t+\phi)=\omega_d\sin(\omega_d t+\phi)=\omega_n\sqrt{1-\zeta^2}\sin(\omega_d t+\phi)$$

此式可以化为

$$\tan(\omega_d t+\phi)=\frac{-\zeta}{\sqrt{1-\zeta^2}}=\tan(\phi)$$

此方程的后面一项与例 12.10 的一样。这里，只有当 $\omega_d t = n\pi$ 或者 $t=\frac{n\pi}{\omega_d}$ 时，才有

$$\tan(\omega_d t+\phi)=\tan(\phi)$$

扫视达到最大角度所需的时间是满足 $t=\frac{n\pi}{\omega_d}$ 的最小 n 值，即 $n=1$。于是

$$T_p=\frac{\pi}{\omega_d}=\frac{\pi}{\omega_n\sqrt{1-\zeta^2}}$$

将 $t=T_p$ 代入式（12.26），可得最大扫视角度为

$$\begin{aligned}\theta(T_p)&=\frac{\gamma}{K}\left[1+\frac{e^{-\zeta\omega_n\frac{\pi}{\omega_n\sqrt{1-\zeta^2}}}}{\sqrt{1-\zeta^2}}\cos\left(\omega_d\frac{\pi}{\omega_n\sqrt{1-\zeta^2}}+\phi\right)\right]\\&=\frac{\gamma}{K}\left[1+\frac{e^{-\zeta\omega_n\frac{\pi}{\omega_n\sqrt{1-\zeta^2}}}}{\sqrt{1-\zeta^2}}\cos(\pi+\phi)\right]\\&=\frac{\gamma}{K}\left[1+\frac{e^{-\zeta\omega_n\frac{\pi}{\omega_n\sqrt{1-\zeta^2}}}}{\sqrt{1-\zeta^2}}\sqrt{1-\zeta^2}\right]\\&=\frac{\gamma}{K}\left(1+e^{-\zeta\frac{\pi}{\sqrt{1-\zeta^2}}}\right)\end{aligned}$$

■

下面利用 Westheimer 模型计算扫视峰速 $\dot{\theta}(t_{pv})$。先计算

$$\begin{aligned}\frac{\partial^2\theta}{\partial t^2}=&\frac{-\gamma e^{-\zeta\omega_n t}}{K\sqrt{1-\zeta^2}}(-\zeta\omega_n\zeta\omega_n\cos(\omega_d t+\phi)+\omega_d\sin(\omega_d t+\phi))\\&+(-\zeta\omega_n\omega_d\sin(\omega_d t+\phi)+\omega_d^2\cos(\omega_d t+\phi))\end{aligned}\tag{12.29}$$

然后，设 $\left.\frac{\partial^2\theta}{\partial t^2}\right|_{t=t_{pv}}=0$，求达到扫视峰速的时间 t_{pv}，可得

$$t_{pv}=\frac{1}{\omega_d}\arctan\left(\frac{\sqrt{1-\zeta^2}}{\zeta}\right)\tag{12.30}$$

将 t_{pv} 代入式（12.27），即可得扫视峰速 $\dot{\theta}(t_{pv})$。根据 Westheimer 模型给定的参数值，设扫视幅度为 $\Delta\theta=\frac{\gamma}{K}$，则由式（12.27），可得

$$\dot{\theta}(t_{pv})=55.02\Delta\theta\tag{12.31}$$

此式表明扫视峰速直接与扫视幅度成正比。但是，图 12.11 所示的实验数据却表明峰速是扫视幅度的指数函数，而不是此 Westheimer 模型所预测的线性函数。式（12.28）的扫视持续时间和式（12.31）的峰速都符合线性系统理论，也就是，用阶跃输入信号激励线性系统时，系统响应达到最大值所需的时间与阶跃输入的幅值大小无关，该时间是一个常数，而响应的最大速度则随着输入的增加而增大。

例 12.12 请从式（12.29）推导出式（12.30）。

解： 将式（12.29）置 0，有

$$\frac{-\gamma e^{-\zeta\omega_n t}}{K\sqrt{1-\zeta^2}}(-\zeta\omega_n\zeta\omega_n\cos(\omega_d t+\phi)+\omega_d\sin(\omega_d t+\phi))$$

$$+(-\zeta\omega_n\omega_d\sin(\omega_d t+\phi)+\omega_d^2\cos(\omega_d t+\phi))=0$$

由于正弦函数外面的乘积项不会等于0，将其去除，可得

$$((\omega_d^2-\zeta^2\omega_n^2)\cos(\omega_d t+\phi))-(2\zeta\omega_n\omega_d\sin(\omega_d t+\phi))=0$$

此式可简化为

$$\tan(\omega_d t+\phi)=\frac{\omega_d^2-\zeta^2\omega_n^2}{2\zeta\omega_n\omega_d}=\frac{\omega_n^2(1-\zeta^2)-\zeta^2\omega_n^2}{2\zeta\omega_n\omega_n\sqrt{1-\zeta^2}}=\frac{1-2\zeta^2}{2\zeta\sqrt{1-\zeta^2}}=\left(1-\frac{1}{2\zeta^2}\right)\tan(\psi)$$

注意，由例12.10可知 $\tan(\psi)=\frac{-\zeta}{\sqrt{1-\zeta^2}}$，并且，由例12.10的图中可知 $\tan(\psi)=\tan(\phi)$。下面先用如下三角恒等式简化上式

$$\tan(\omega_d t+\phi)=\frac{\tan(\omega_d t)+\tan(\phi)}{1-\tan(\omega_d t)\tan(\phi)}$$

可得

$$\tan(\omega_d t)+\tan(\phi)=(1-\tan(\omega_d t)\tan(\phi))\times\left(1-\frac{1}{2\zeta^2}\right)\tan(\psi)$$

$$=\tan(\psi)-\frac{1}{2\zeta^2}\tan(\psi)-\left(1-\frac{1}{2\zeta^2}\right)\tan(\psi)\tan(\omega_d t)\tan(\phi)$$

再代入 $\tan(\psi)=\tan(\phi)$ 并重排，得到

$$\tan(\omega_d t)\left[1+(\tan(\phi))^2\left(1-\frac{1}{2\zeta^2}\right)\right]=\frac{-\tan(\phi)}{2\zeta^2}$$

即

$$\tan(\omega_d t)=\frac{-\tan(\phi)}{2\zeta^2\left(1+(\tan(\phi))^2\left(\frac{2\zeta^2-1}{2\zeta^2}\right)\right)}=\frac{-\tan(\phi)}{2\zeta^2+(\tan(\phi))^2(2\zeta^2-1)}$$

再利用 $\tan(\phi)=\frac{-\zeta}{\sqrt{1-\zeta^2}}$，上式可以简化为

$$\tan(\omega_d t)=\frac{\frac{\zeta}{\sqrt{1-\zeta^2}}}{2\zeta^2+\frac{\zeta^2(2\zeta^2-1)}{1-\zeta^2}}=\frac{\sqrt{1-\zeta^2}}{\zeta}$$

即

$$t_{pv}=\frac{1}{\omega_d}\arctan\left(\frac{\sqrt{1-\zeta^2}}{\zeta}\right)$$ ■

Westheimer注意到了模型的扫视持续时间一幅度以及峰速一幅度这两个曲线与实验数据都有差别，并且从实验数据的峰速一幅度曲线的非线性，可以推测扫视眼动系统并不是线性的。他也注意到扫视系统的输入不是突变的阶跃函数。不过，无论如何，这个模型与20°扫视的眼动位置实验数据之间具有很好的拟合，但对其他幅度的扫视并非如此。有趣的是，如果改用下一节所述的其他输入激励函数，而不是阶跃函数，此Westheimer二阶模型就可以模拟任意幅度的扫视。由于Westheimer眼动模型很简单，因此，至今仍很常用。

12.5　扫视控制器

生理系统建模中一个难以解决的问题是缺乏系统输入激励信号的详细信息。例如，快速眼动系统的输入是大脑中枢神经系统传导到眼部肌肉的神经信号。这个输入信号由成千上万个神经元的高速动作电位发放构成，现在还无法获得这种输入信号的详细记录。记录这种信号要用特殊的介入式技术和仪器，目前还不能实现。因此，与很多其他生理系统一样，建立眼动系统模型时所遇到的难题就是不知道输入信号。不过，我们可以利用下面所介绍的间接方法获得有关快速眼动系统的输入信息。

1964 年，Robinson 设计了一个实验，用于测量扫视过程中眼球的输入激励信号。他用隐形眼镜吸住一只眼球，使其固定不动，并让另一只眼球在两个目标之间进行扫视运动。由于扫视过程中传导到两个眼球的神经控制信号相同，Robinson 推断，加在固定不动的眼球上的传感器所记录到的输入信号，就是驱动另一只眼球的输入信号。如图 12.13 所示，他指出，扫视过程中驱动眼球的肌张力信号是一个脉冲信号加上一个阶跃信号，即脉冲－阶跃输入信号。

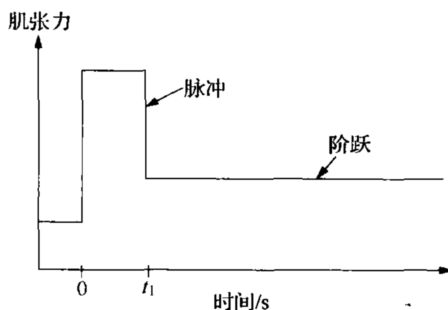


图 12.13　扫视期间记录到的肌张力变化示意图

有人用微电极记录动眼神经元的电活动。图 12.14 所示就是用玻璃管微电极记录动眼神经核电信号的实验，动眼神经核是驱动扫视的主要神经元集群。自 Robinson 的开创性工作之后，人们开发了很多通过测量眼肌来研究扫视控制器的其他实验。例如，1975 年，Collins 和同事们利用微型“C”形测力传感器，于肌腱部位在体检测了人眼无约束活动时的肌张力。这种方法可以较好地测量每块眼肌所产生的独立肌张力，而不是图 12.13 所示的两块相互拮抗的眼肌的作用力之和。

图 12.14 中所记录的跨膜锋电位就是本书第 11 章所述的动作电位发放。在眼球开始运动前 5ms 左右，控制扫视的神经元就开始爆发式的动作电位发放，并持续到眼球几乎到达目标位置为止。图中显示了从角度 θ_0 到 θ_3 的眼球相对位置。开始时眼球处于 θ_0 位置，此时眼肌的输入控制为 0，眼肌处于完全放松状态。为了将眼球从 θ_0 移动到 θ_1，神经元产生爆发式的脉冲发放串。然后，为了使眼球保持在 θ_1 位置，神经元产生有规律的平稳发放，其发放率正比于眼肌收缩的量。接下来是第二次扫视，眼球从 θ_1 移到 θ_2，这次扫视的速度比第一次慢，但是持续时间却大致相同。不同的是，驱动眼球运动的神经元发放数目较少，这就意味着输入信号的幅值比第一次扫视要小。在保持 θ_2 位置期间，由于眼肌比在 θ_1 位置时收缩得更短，因此，神经元平稳发放的速率更高。第三次扫视，眼球反向移动到 θ_3。此时眼肌舒张伸长，其输入为 0，也就是，没有动作电位刺激肌肉。在 θ_3 位置期间，由于眼肌长度比 θ_1 位置时更

长，因此，神经元平稳发放水平比 θ_1 时低。

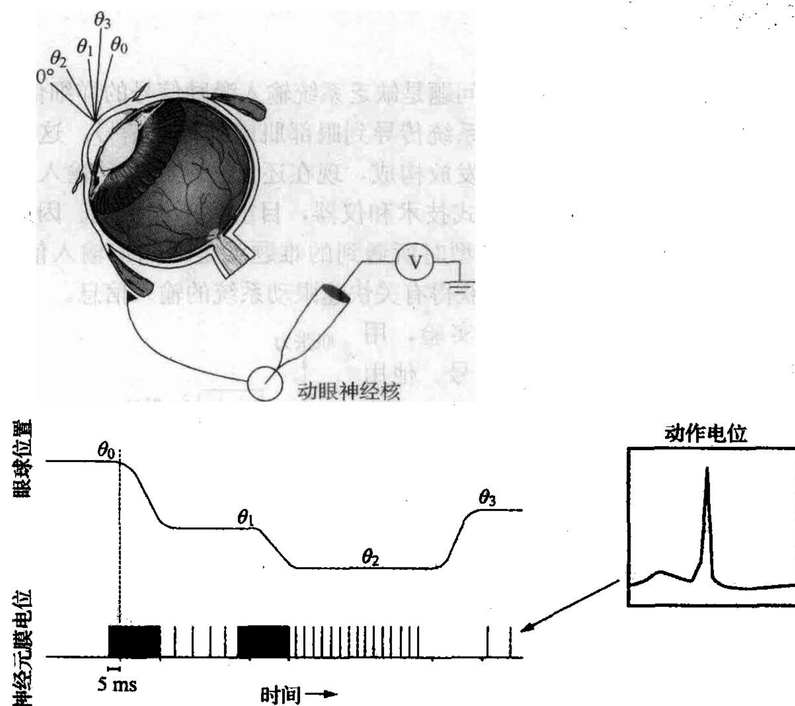


图 12.14　连续扫视过程中玻璃微电极记录单个神经元电活动的示意图

这里，必须区分两个概念，一是肌张力，也就是眼肌所产生的力；另一是激活态张力发生器（active-state tension generator），是指眼肌内部的力发生器，它所产生的内力，通过眼肌内部各种因素的转化作用再变成肌张力。肌张力是外在的，可以测量，而激活态张力是内在的，不可测量。激活态张力几乎与眼肌的神经输入信号相吻合。图 12.14 所示的神经活动模式表明：

1）收缩的眼肌（即主动肌）先受到一个脉冲式的刺激信号，接着是保持眼球位置的一个阶跃式信号。

2）舒张伸长的眼肌（即拮抗肌）在扫视过程中不接受刺激，刺激信号中断，换句话说，可以看做受到一个负向置零的脉冲信号。然后，紧随其后的还是保持眼球位置的阶跃信号。

图 12.15 分别显示了主动肌神经输入信号 N_{ag} 和拮抗肌神经输入信号 N_{ant} 的量化曲线。其中，脉冲输入用于使眼球尽快移动到目标位置，而阶跃输入则使眼球保留在特定位置。

前人的研究结果表明激活态张力曲线与神经控制信号曲线并不一样，前者可以看做是脉冲一阶跃神经信号的低通滤波。图 12.15 虚线所示是激活态张力，它具有 τ_{ac} 和 τ_{de} 两个时间常数。这种低通滤波作用被认为与 Ca^{2+} 离子的活动有关。有人提

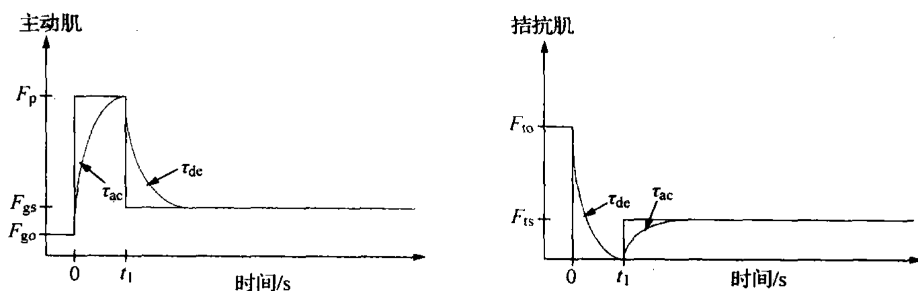


图 12.15　主动肌和拮抗肌的输入控制信号 N_{ag}和 N_{ant}（实线），及其相应的激活态张力 F_{ag}和 F_{ant}（虚线）。

注：激活时间常数 τ_{ac}与去激活时间常数 τ_{de}是不同的。时间间隔 t_1为脉冲信号的宽度。

出，主动肌和拮抗肌具有不同形式的时间常数，有些人认为主动肌活动的时间常数与神经元发放频率相关，还有人认为主动肌活动的时间常数是扫视幅度的函数。本书为了简化，假设描述主动肌和拮抗肌活动的激活时间常数和去激活时间常数相等。

1964 年，Robinson 根据实验结果，提出了一种描述快速眼动的模型。通过改变脉冲一阶跃输入信号的幅值大小，来仿真 5°～40°范围内的眼球扫视运动。由这些仿真结果建立的扫视位置随时间变化的曲线与生理实验结果一致。但是，速度一时间曲线却与生理表现不同。在修正此模型缺陷的过程中，20 世纪 60 年代至 90 年代期间，人们通过各种眼动系统的生理实验和建模研究，开发了与生理系统更接近的模型。这些工作的关键是建立各种动眼肌模型。

12.6　动眼肌模型的建立

水平快速眼动系统由内直肌和外直肌这对眼肌驱动，显然，要建立该系统的模型，必须有精确的描述眼肌活动的模型。实际上，前述 Westheimer 模型并没有包含眼肌模型，只包括眼球的惯性、眼球与眼眶之间的摩擦力、视神经等附件所产生的粘弹性。本节将介绍动眼系统肌肉模型的建立过程。其实，研究肌肉模型所涉及的内容非常广泛，从描述肌小节的微观模型，到描述多细胞组合的宏观模型。其中，细胞群的宏观模型用具有集中参数的普通力学元件来表示。这里我们将介绍根据生理实验测量数据建立的一种动眼肌宏观模型。模型元件包括：一个激活态张力发生器（代表输入信号）、几个弹性元件和粘弹性元件。下面的每一小节单独介绍一个元件，最终阐述整个模型。注意，本节末尾最后给出的线性眼肌模型将彻底修正本节前面部分所述的模型，这些前面部分阐述的是模型发展的历史过程，有利于读者理解目前所用的眼肌模型。

12.6.1　被动弹性特性

肌肉的被动弹性特性可以通过实验测定，方法是拉伸处于非激活状态的肌肉，

并记录其张力。图 12.16 所示的数据曲线就是一个典型的眼直肌张力记录曲线。可见，拉伸肌肉所需的张力大小是拉伸距离的非线性函数。因此，为了精确建立其模型，必须使用非线性弹簧元件。注意，张力为 0 时的长度增量是指眼睛处于平直注视正前方的第一眼位时的眼肌长度变化，此时，眼肌大约被拉伸3mm，眼肌总长度约为 37mm。

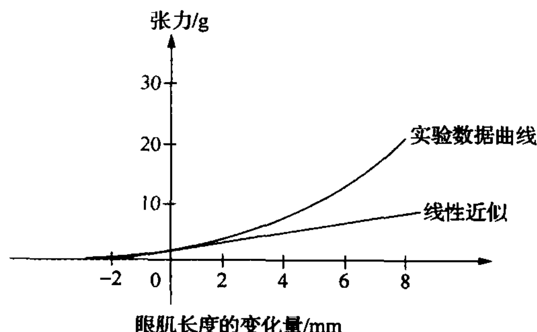


图 12.16　未激活眼肌的张力－长度变化曲线示意图

注：其线性近似直线的斜率表示肌肉的被动弹性系数 K_{pe}。

图 12.16 所示的实验数据曲线必须在某个工作点附近进行线性化，这样才能用于肌肉的线性模型。工作点应该取在弹簧元件工作范围的中点。图 12.16 中的近似直线相切于实验曲线上眼球的第一眼位处，是该区域弹簧行为的线性逼近。设眼球半径为 11mm，为了便于分析，先计算以下数据：

$$1\text{gm}=9.806\times10^{-3}\text{N}$$

$$1^{\circ}=0.192\text{mm}=1.92\times10^{-4}\text{m}$$

线性近似直线的斜率 K_{pe} 约为

$$K_{pe}=0.2\,\frac{1\text{g}}{1^{\circ}}\times\frac{9.806\times10^{-3}\text{N}}{1\text{g}}\times\frac{1^{\circ}}{1.92\times10^{-4}\text{m}}=10.2\,\frac{\text{N}}{\text{m}}$$

它表示被动弹性元件的弹性系数。工作区域的选择对于曲线斜率的确定至关重要。此处，用了早期提出的眼直肌工作区域中的一个位置。在多数有关眼动系统的论文中，这个 K_{pe} 一般被减去，不会用到。我们会在本章后面修正工作点位置，彻底除去模型中的这个 K_{pe} 元件。

12.6.2　激活态张力发生器

一般，肌肉所产生力的大小与所受到的刺激成正比。产生力的元件就是激活态张力发生器。注意，激活态张力是指肌肉内部产生的作用力，不要与肌张力相混淆。如图 12.17 所示，激活态张力发生器与被动弹性元件一起构成肌肉模型。肌张力 T、激活态张力 F 和弹性系数 K_{pe} 之间的关系为

$$T=F-K_{pe}x \qquad (12.32)$$

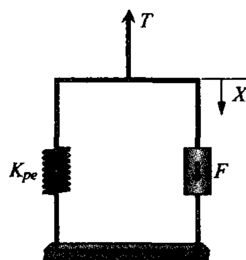


图 12.17　一种肌肉模型示意图

注：其中包括一个激活态张力发生器 F 和一个被动弹性元件 K_{pe}。一旦受到激活态张力发生器的刺激，肌肉就会产生张力 T。

多年来，为了计算各个不同输入刺激水平下眼肌激活态张力的数值，研究人员进行了许多人体眼部肌肉的等长收缩实验。这种实验一般结合

斜视矫正手术进行。手术时，肌肉与眼球的连接先被切断，然后再重新连到眼球上，以纠正斜视眼。如图 12.18 所示，受到刺激的肌肉所产生的张力是肌肉长度的函数。这组数据来自同一只眼球上断开的眼外直肌，当时，另一只没有手术的眼球从鼻侧 N 到颞侧 T，在−45°～45°之间分别固定注视于不同的位置。该实验设计的依据是两只眼球的肌肉所接收的神经输入信号相同，即 Hering 相同神经支配假设，因此，自由活动的那只眼球外直肌的激活态张力应该与另一只眼球上断开的外直肌相同。在活动眼球的每个固定注视位置下，将另一只眼球断开的外直肌拉至相应的位置，同时测量其张力，这样就形成了图中所示的各点数据。其中，粗线就表示眼球正常转动时外直肌在各个位置下的肌张力。45°T 曲线是输入刺激为 0 情况下的测量数据，表示肌肉的被动力学特性。注意，肌肉所产生的张力是其长度的非线性函数。

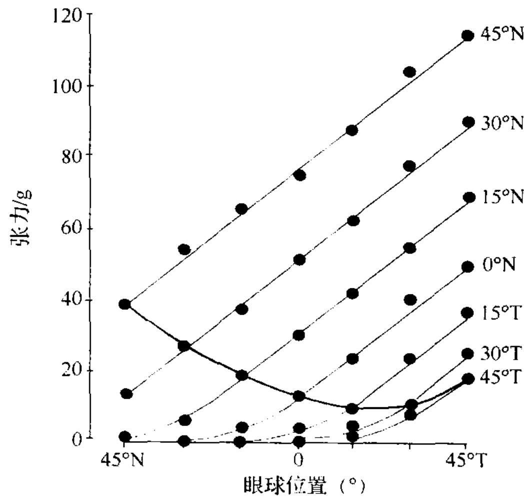


图 12.18 不同激活水平下外直肌的长度−张力曲线

注：圆点表示斜视矫正手术中从断开的外直肌上记录的张力数据。测量时，另一只没有手术的眼球分别固定注视−45°～45°范围之间的目标（摘自 Collins、O'Meara 和 Scott，1975）。

对比图 12.17 的模型与图 12.18 的数据可知，如果将 30°T 到 45°N 的每条曲线减去被动弹性部分（即 45°T 曲线所表示的数值），剩余的就应该是激活态张力。图 12.19 左图所示的就是 15°N 曲线（实线）及其相应的激活态张力（虚线）。图 12.18 的其他曲线的计算结果相似，为了图示的简洁明了，这里省略不显示。虚线表示的激活态张力是肌肉长度的函数。如果这条减下来的曲线只包含纯粹的激活态张力元件的作用，那么它应该是一条与输入大小相对应的水平直线，图 12.19 右图就显示了这样一条激活态张力水平直线。由此可见，要么激活态张力发生器是一个非线性元件，也就是，此模型可能缺少某些非线性或线性元件；要么建立此模型所用的某些假设前提可能是错误的。这里，暂时假定所用的前提是正确的，只是模型缺少了元件。本节后面会修正这个问题。

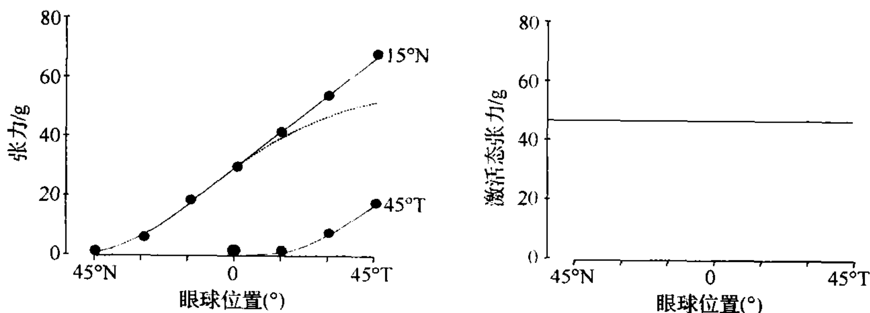


图 12.19 左图：眼外直肌在 15°N 和 45°T 两个激活位置水平下的长度－张力曲线，圆点表示斜视矫正手术中从断开的外直肌上记录的张力数据，测量时，另一只没有手术的眼球固定注视于特定位置上的目标，虚线是 15°N 曲线减去 45°T 曲线之后的结果，可见它所表示的激活态张力仍然是眼球位置的函数；右图：激活态张力－眼球位置之间的理论关系（摘自 Collins、O'Meara 和 Scott，1975）

12.6.3 弹性

如图 12.19 所示，虽然从张力数据中减去了被动弹性部分，但是如前所述，肌肉的激活态张力与长度之间仍然存在一定的关系。为了获得两者之间的这种关系，如图 12.20 所示，在模型中新添加一个弹性元件 K，并用以下方程描述模型

$$T=F-K_{pe}x-K_x \tag{12.33}$$

新元件 K 决定图 12.19 左边虚线表示的减下来的曲线的斜率。眼球处于第一眼位时所对应的 K 值约为 0.8g/(°) =40.86N/m（这是文献所用的值）。用同样的方法可以计算图 12.18 所示其他曲线在第一眼位处的斜率 K，其结果都近似等于此 15°N 曲线的斜率值。至此，需要进行其他实验研究来进一步完善该眼肌模型。

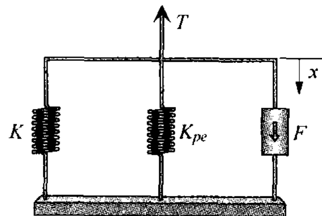


图 12.20 一种眼肌模型示意图

注：其中包括激活态张力发生器 F、被动弹性元件 K_{pe} 和弹性元件 K。一旦受到激活态张力发生器的刺激，肌肉就会产生张力 T。

串联弹性元件

Levin 和 Wyman（1927）以及 Collins（1975）的实验结果都表明眼肌模型中需要增加串联弹性元件。图 12.21 显示了他们的实验设计以及典型数据。该实验被称为快速释放实验，其步骤如下：在肌肉上悬挂一重物；从 t_1 时刻开始给肌肉施加充分的电刺激；再于 t_2 时刻将重物释放。这样，在 t_2 时刻重物释放时，肌肉几乎瞬间改变了长度。能够瞬间改变自身长度的元件只有弹簧。因此，为了体现眼肌的这种行为，在激活态张力元件上串联一个弹簧 K_{se}，这个弹簧就称为串联弹性元件。图 12.22 显示了这种改进眼肌模型。

根据 Collins 在 1975 年所做的眼直肌实验，可估计 K_{se} 的值约为 125N/m（即 2.5gm/(°)）。由于 K_{se} 的值与上述 K 值不相等，还需要在模型中再添加一个弹性元件。

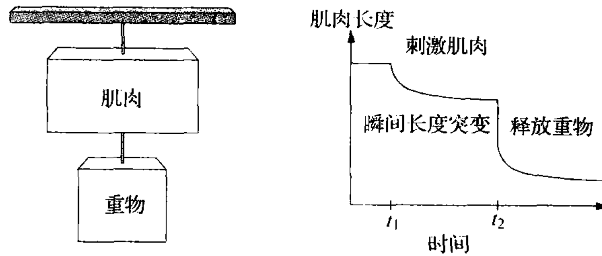


图 12.21 快速释放实验示意图

注：左图：实验的物理装置；右图：实验记录的典型数据。t_1时刻肌肉开始受到充分的电刺激，t_2时刻重物释放。

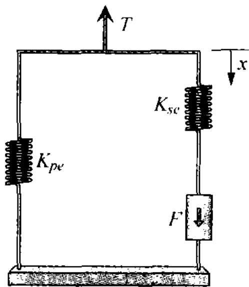


图 12.22 一种眼肌模型示意图

注：其中包括激活态张力发生器 F、被动弹性元件 K_{pe} 和串联弹性元件 K_{se}。一旦受到激活态张力发生器的刺激，肌肉就会产生张力 T。

长度一张力弹性元件

由于 K_{se}与 K 不相等，如图 12.23 左边所示，模型中再加一个与激活态张力元件并联的弹性元件 K_{lt}，该元件称为长度一张力弹性元件。为了便于分析，应用图 12.19 所示的方法将 K_{pe}减去，得到图 12.23 右边所示的模型。下面根据此模型，推导 K_{lt}的计算公式。首先，分别求图中节点 1 和节点 2 的作用力之和，即

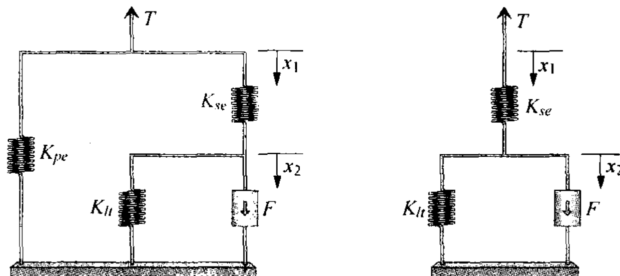


图 12.23 左图：一种眼肌模型，其中激活态张力发生器 F 与长度一张力弹性元件 K_{lt} 并联，再与串联弹性元件 K_{se} 串联，最后，这些元件与被动弹性元件 K_{pe} 并联，一旦受到激活态张力发生器的刺激，肌肉就会产生张力 T；右图：眼肌模型的其他部分与左图相同，只是去除了 K_{pe}

$$T=K_{se}(x_2-x_1) \tag{12.34}$$

$$F=K_{lt}x_2+K_{se}(x_2-x_1)\rightarrow x_2=\frac{F+K_{se}x_1}{K_{se}+K_{lt}} \tag{12.35}$$

将 x_2代入式（12.34），得到

$$T=\frac{K_{se}}{K_{se}+K_{lt}}(F+K_{se}x_1)-K_{se}x_1=\frac{K_{se}}{K_{se}+K_{lt}}F-\frac{K_{se}K_{lt}}{K_{se}+K_{lt}}x_1 \tag{12.36}$$

此式是一个直线方程，其 y 轴截距为$\frac{K_{se}}{K_{se}+K_{lt}}$，斜率为$\frac{K_{se}K_{lt}}{K_{se}+K_{lt}}$。图 12.19 所示长度一张力曲线的斜率为 $K=0.8\text{g}/(°)=40.86\text{N/m}$，因此

$$K=\frac{K_{se}K_{lt}}{K_{se}+K_{lt}}=40.86\ \frac{\text{N}}{\text{m}} \tag{13.37}$$

求解此方程，可得 K_{lt}为

$$K_{lt}=\frac{K_{se}K}{K_{se}-K}=60.7\ \frac{\text{N}}{\text{m}} \tag{13.38}$$

12.6.4 力—速度关系

早期实验表明，肌肉不仅具有弹性，还具有粘性。图 12.24 所示的作用力恒定的等张收缩实验可以用于研究肌肉的粘性。实验时，肌肉和负载分别加在一根杠杆的两端，其杠杆比很大。杠杆使负载加在肌肉上的重力（即质量×比重）变为负载重力除以杠杆比，而负载加在肌肉上的惯性力（即质量×加速度）则为负载惯性力除以杠杆比的平方。因此，在这种实验条件下，肌肉等张收缩期间，负载施加在肌肉上的惯性力可以忽略不计。另一个假设是，如果杠杆比没有使负载惯性力的作用减小到可以忽略的程度，那么，还可以在最大速度处进行测量。此时加速度为 0，惯性力也就等于 0。如果这两个假设成立，那么，不同负载重力作用下所测得的实验数据就不受惯性力的影响。

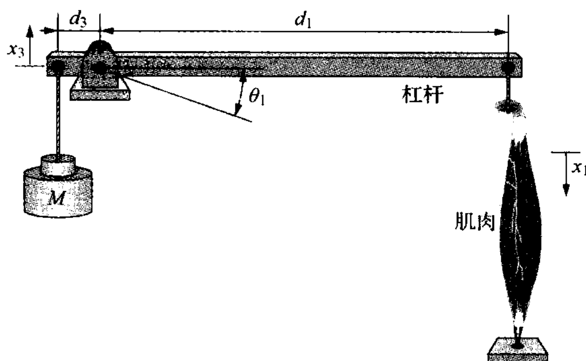


图 12.24 等张实验示意图

注：负载和肌肉分别挂在杠杆两端，肌肉的另一端固定。实验时，肌肉伸缩到最佳长度。

在这种条件下，等张实验中肌肉会伸长到其最佳长度。实验时，加上负载 M，同时给予肌肉一定强度的电刺激输入激励，记录肌肉伸缩到达的位置。图 12.25 所示两条曲线分别记录了一大一小两种负载作用下的等张实验过程中肌肉伸缩的数据。注意，两个响应的持续时间大致相等，与负载大小无关，只是负载大的时候，起始处的时间延时（即潜伏期）明显长得多。还有，负载越大，肌肉的收缩量就越小。最大伸缩速度可以从所记录的肌肉伸缩位置数据中求得。为了估计肌肉的粘性系数，

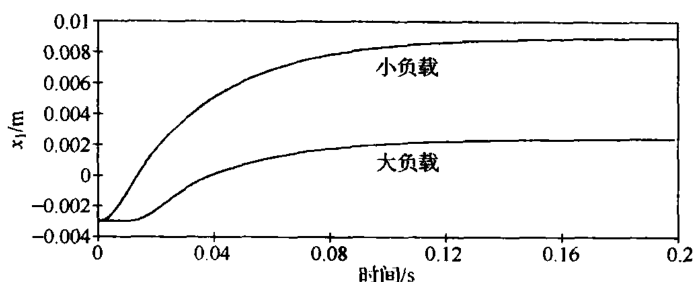


图 12.25　一大一小两种负载作用下肌肉的响应曲线

在同样的电刺激水平下，用多个不同的负载重复进行实验，然后计算出每次实验的最大速度。图 12.26 显示了负载比 P/P_0 与最大速度之间的典型关系曲线，其中，$P=Mg$，P_0是肌肉受到最大刺激时的等长收缩张力（即肌肉可以提起的最大重量）。该曲线通常称为力一速度曲线。

显然，力一速度曲线是非线性的，呈现双曲线形状。如果用小于最大刺激的不同强度电刺激兴奋肌肉，那么，就可以得到图 12.27 所示的对应于不同激活态张力的一组力一速度曲线，每条曲线上标出的数值就是激活态张力的大小。这些力一速度曲线的特性与图 12.26 所示曲线很相似。注意，在眼肌的正常工作范围内（即 800°/s 左右），激活态张力较小时的力一速度曲线的斜率与激活态张力较大时的情况差别很大。

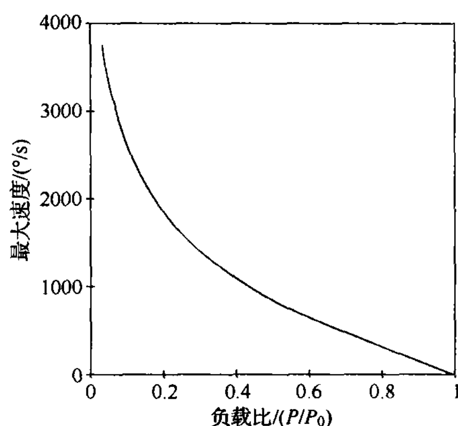


图 12.26　力一速度曲线示意图

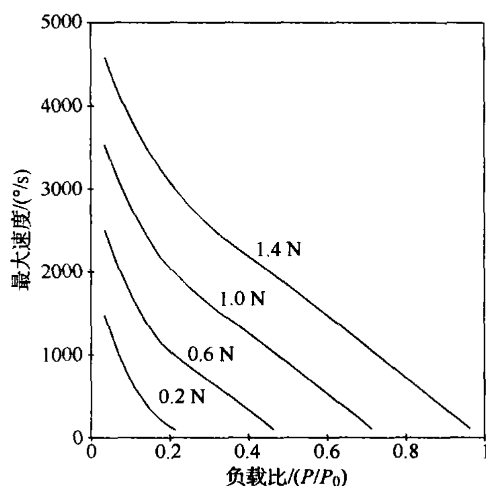


图 12.27　激活态张力从 0.2N 变化到 1.4N 时测得的一组力一速度曲线

下面将上述等张实验所得到的粘性作用添加到模型中。如图 12.28 所示，在激活态张力发生器和长度一张力弹性元件上再并联一个粘性元件。下面的例 12.13 将

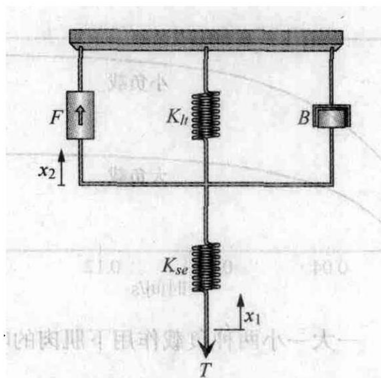
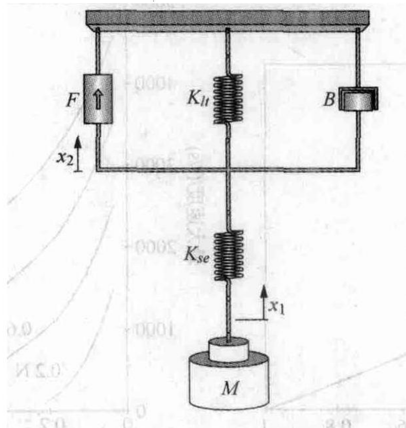


图 12.28 一种眼肌模型示意图

注：其中激活态张力发生器 F 与长度一张力弹性元件 K_{lt} 以及粘性元件 B 并联，再与弹性元件 K_{se} 串联。为了简化，被动弹性元件 K_{pe} 已去掉。一旦受到激活态张力发生器的刺激，肌肉就会产生张力 T。

应用该模型仿真等张实验，并分析模型的行为以及新添加粘性元件的作用。现在暂时假设此粘性元件是线性的。为了简化，杠杆随同加速度项 $M\ddot{x}_1$ 被去除。本章后面会考虑杠杆，并作进一步分析。同样为了简化，例 12.13 也将被动弹性元件 K_{pe} 减除。

例 12.13 下图所示系统是等张实验的一种模型。假设加速度项 $M\ddot{x}_1$ 可以忽略不计。请计算最大速度随负载变化的函数，并作出其曲线图。



解：假设 $\dot{x}_2 > \dot{x}_1$，且 $x_1 > 0$。令 $K_{st} = K_{se} + K_{lt}$。图中节点 1 和节点 2 上的作用力之和分别为

$$Mg = K_{se}(x_2 - x_1) \rightarrow x_1 = x_2 - \frac{Mg}{K_{se}}$$

$$F = B\dot{x}_2 + K_{lt}x_2 + K_{se}(x_2 - x_1)$$

将 x_1 代入第二个方程，可得

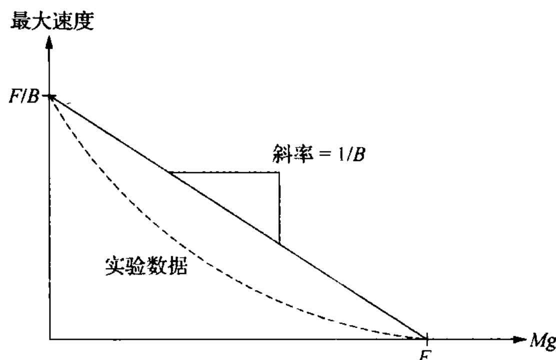
$$F=B\dot{x}_2+K_{lt}x_2+Mg$$

求解此方程，可得

$$x_2(t)=\frac{F-Mg}{K_{lt}}\left(1-\mathrm{e}^{-\frac{K_{lt}t}{B}}\right)$$

$$\dot{x}_2(t)=\frac{F-Mg}{B}\mathrm{e}^{-\frac{K_{lt}t}{B}}$$

各种负载下的最大速率为 $V_{\max}=\frac{F-Mg}{B}$，并且，由于 $\dot{x}_1=\frac{\mathrm{d}}{\mathrm{d}t}\left(x_2-\frac{Mg}{K_{se}}\right)$，因此 $\dot{x}_1=\dot{x}_2$。下图描述了最大速率与负载之间的线性关系，以及对应的实验数据曲线。



■

粘性元件的线性假设似乎有问题，以上例 12.13 的分析结果给出的是负载与最大速度之间的线性关系，而图 12.26 所示的等张实验数据显然是非线性的。因此，粘性元件应该是非线性的。

通常，肌肉粘性用非线性 Hill 双曲线来表征，即

$$V_{\max}(P+a)=b(P_0-P) \tag{12.39}$$

式中 $V_{\max}$——最大速度；

P——外加作用力，就是负载重量 Mg；

P_0——等长张力，是指肌肉可以提起的最大重量；

a 和 b——双曲线渐近线的经验常数。

由 Hill 数据可知

$$a=\frac{P_0}{4} \text{ 且 } b=\frac{V_{\max}}{4}$$

于是，将此 a 值和 b 值代入式（12.39）之后，Hill 方程变为

$$P=P_0-\frac{V_{\max}\ (P_0+a)}{b+V_{\max}}=P_0-BV_{\max} \tag{12.40}$$

其中

$$B=\frac{P_0+a}{b+V_{\max}} \tag{12.41}$$

B 表示元件的粘性。显然，由于上式分母中存在速度项 $V_{\max}$，因此，粘性所产生的作用力是非线性的。

对于眼动模型，根据图12.28的定义，用$\dot{x}_2$代替V_{max}，用肌肉张力T代替P，用激活态张力F代替P_0，则式（12.40）和式（12.41）可以重写为

$$T=F-B\dot{x}_2 \tag{12.42}$$

和

$$B=\frac{F+a}{b+\dot{x}_2} \tag{12.43}$$

例12.14 请写出图12.28所示眼肌模型产生张力的方程。

解：假设$\dot{x}_2>\dot{x}_1$，则节点1和节点2上的作用力之和为

$$T=K_{se}(x_2-x_1)\rightarrow x_2=\frac{T}{K_{se}}+x_1$$

$$F=B\dot{x}_2+K_{lt}x_2+K_{se}(x_2-x_1)$$

将x_2代入第二个方程，消去其中的x_2，得到

$$F=B\dot{x}_2+(K_{se}+K_{lt})\left(\frac{T}{K_{se}}+x_1\right)-K_{se}x_1$$

可得T的解为

$$T=\frac{K_{se}}{K_{se}+K_{lt}}F-\frac{K_{se}K_{lt}}{K_{se}+K_{lt}}x_1-\frac{K_{se}B}{K_{se}+K_{lt}}\dot{x}_2$$

将式（12.43）的B代入上式，就得到动眼肌的非线性模型。 ■

已有研究人员根据肌肉拉伸和收缩两种不同情况，求得Hill方程中的a值和b值。结果表明，肌肉拉伸与收缩这两个动力学过程是不同的。不过，肌肉收缩和拉伸的粘性作用的方程都可以由式（12.43）给出，只是a和b的参数值不同。例如，Hsu及其合作者将动眼肌收缩和拉伸的粘性分别表示为

$$B_{ag}=\frac{F_{ag}+AG_a}{\dot{x}_2+AG_b} \tag{12.44}$$

$$B_{ant}=\frac{F_{ant}-ANT_a}{\dot{x}_2+ANT_b} \tag{12.45}$$

式中，AG_a、AG_b、ANT_a和ANT_b分别是根据肌肉收缩（即主动肌）和拉伸（即拮抗肌）的渐近线得到的参数。

式（12.44）和式（12.45）定义的B_{ag}和B_{ant}为速度的非线性函数。为了建立线性肌模型，可以将B_{ag}和B_{ant}进行线性化处理，也就是，作出图12.27所示双曲线函数的近似拟合直线。拮抗肌的激活水平一般在5%（即$F=0.2$N），而主动肌激活水平一般在100%（即$F=1.4$N）。

12.7 线性眼肌模型

本节将重新讨论线性眼肌模型的静态和动态特性。图12.29所示是改进的线性动眼肌模型。模型的每个元件都是根据生理实验数据建立的线性元件，包括粘性元

件 B_2 与串联弹性元件 K_{se} 的并联组合，再加上激活态张力发生器 F、粘性元件 B_1 和长度一张力弹性元件 K_{lt} 的并联组合。变量 x_1 和 x_2 表示模型中各个刚性元件偏离平衡位置的位移。该模型与前述动眼肌模型在结构上的惟一区别就是添加了粘性元件 B_2，也去除了被动弹性元件 K_{pe}。正如下面将介绍的，粘性元件 B_2 是描述肌肉非线性力一速度特性的关键因素，而被动弹性元件 K_{pe} 是不必要的。

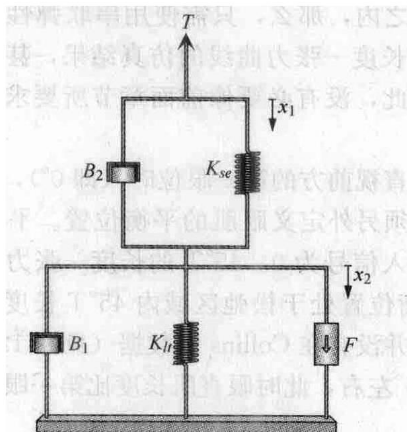


图 12.29　改进的线性眼肌模型示意图

注：其中包括激活态张力发生器 F、长度—张力弹性元件 K_{lt} 和粘性元件 B_1 三者的并联组合，再连接于串联弹性元件 K_{se} 和粘性元件 B_2 两者的并联组合。一旦受到激活态张力发生器的刺激，肌肉就会产生张力 T。

线性动眼肌系统中需要两种弹性元件，这是根据生理数据得出的结论。如前所述，串联弹性元件 K_{se} 及其数值是根据 Collins 的等张—等长收缩快速释放实验设定的。长度—张力弹性元件 K_{lt} 则从长度—张力曲线的斜率得到。两个线性粘性元件 B_1 和 B_2 的使用则基于等张实验所得数据，其值可以由本节下面所述的仿真结果得到。

12.7.1　长度—张力曲线

由图 12.18 可见，当张力小于 10g 时，激活和非激活肌肉的长度—张力关系曲线是非线性的，非线性弹性假设就是根据这一实验结果提出的。1975 年，Collins 利用一种微型“C”形测力传感器，在人眼无约束活动过程中，于肌腱部位在体测量了肌张力。图 12.18 所示数据是手术过程中左眼直肌上测得的数据（参见 12.6.2 节）。测量时，被测者根据要求，用自由活动的右眼注视 $-45°$ 颞侧 T 到 $45°$ 鼻侧 N 上分布的各个目标物，从而诱发不同的神经控制刺激水平。同时，在左眼直肌的不同肌长度下（该眼直肌与眼球的连接已被断开），记录其对应于眼球位置从 $-45°$ 变化到 $45°$ 时的等长收缩张力。实验中整个眼球位置的变化范围所引起的肌肉长度总变化量约为 18mm。Collins 将长度—张力曲线描述为“张力大于 10g 时是平行直线，小于 10g 时动眼肌开始松弛。”他还指出，在使用力传感器进行在体检测时，各种眼动期间眼直肌张力的正常值不会小于 10g，也就是肌肉不会进入松弛状态。

建立动眼系统肌肉模型时，必须使模型能够准确再现眼直肌正常工作范围内的静态特性。因此，动眼肌模型具有的长度－张力特性都必须包含大于10g张力的各条平行直线。前面章节所建立的眼肌模型考虑了张力在10g以下的长度－张力曲线的非线性。既然动眼肌不会工作于10g以下的张力状态，那么，即使线性模型的行为与这部分非线性区域不匹配，也没什么关系。正如本节将要证明的，如果将模型限定在动眼肌的工作范围之内，那么，只需使用串联弹性元件以及长度—张力弹性元件，就可以获得准确的长度—张力曲线的仿真结果，甚至在激活态张力为0的情况下也可以正确仿真。因此，没有必要像前面章节所要求的那样，在眼肌模型中使用被动弹性元件。

如前所述，眼睛处于直视前方的第一眼位时（即0°），动眼系统的眼直肌并不是处于平衡状态，因此，必须另外定义眼肌的平衡位置。平衡是指肌张力为0时的无伸缩状态，此时肌肉的输入信号为0。45°T的长度—张力曲线被设为激活态张力为0。一般，眼外直肌的平衡位置处于松弛区域内45°T长度—张力曲线与水平轴相交的地方。注意，这个交点并没有在Collins的数据（见图12.18）中显示出来，不过，根据报道，该交点位于15°左右，此时眼直肌长度比第一眼位时大约短3mm，该数值是文献中常用的值。

由于正常眼动期间，眼肌并不会处于松弛状态，因此，从眼肌工作区域计算得到的平衡位点是更切合实际的。这里，我们根据松弛区域以上45°T长度—张力曲线的近似拟合直线来确定平衡位点。该拟合直线与水平轴的交点所给出的平衡位点值约为－19.3°。使用－19.3°平衡位点之后，就没有必要像其他模型那样，再用弹性元件K_{pe}表示无刺激状态下眼肌的被动弹性特性。

图12.29所示线性眼肌模型所产生的张力可以表示为

$$T=\frac{K_{se}}{K_{se}+K_{lt}}F-\frac{K_{se}K_{lt}}{K_{se}+K_{lt}}x_1 \tag{12.46}$$

已知眼肌的长度—张力曲线的斜率等于0.8g/(°)＝40.86N/m，且$K_{se}=2.5$g/°＝125N/m，因上式的斜率为

$$\frac{K_{se}K_{lt}}{K_{se}+K_{lt}} \tag{12.47}$$

就可以求得K_{lt}的值为1.2g/(°)＝60.7N/m。

为了计算眼球在各个不同注视位置下的激活态张力，可以从式（12.46）求解F，得到各个神经控制水平下的拟合直线，即

当$\theta>0°$时（即鼻侧N方向），有 $F=0.4+0.0175\theta$N (12.48)

当$\theta\leqslant 0°$时（即颞侧T方向），有 $F=0.4+0.012\theta$N (12.49)

式中 θ——眼球偏离第一眼位的角度，单位为°，且$\theta=5208.7\times(x_1-3.705)$。

注意，$5208.7=\frac{180}{\pi r}$，其中r为眼球半径。

图12.30所示是利用式（12.46）～式（12.49）求得的一组眼肌静态长度—张力

曲线，可用于描述长度—张力的实验结果。这里没有考虑肌肉在松弛区域的活动情况，因为眼直肌正常工作时不会进入该区域。在眼肌工作区域内，此图所示的长度—张力曲线与图 12.18 所示的实验数据吻合得非常好。

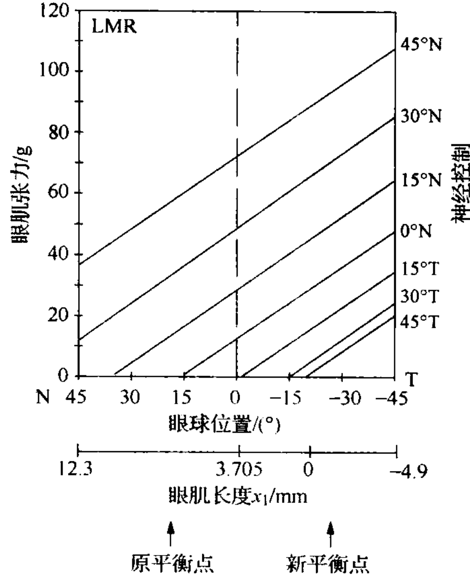


图 12.30　利用线性眼肌模型，即式（12.46）～式（12.49）求得的长度—张力曲线

注：其中，输入信号大小分别为：45°N 时，$F=130$g；30°N 时，$F=94.3$g；15°N 时，$F=64.9$g；0°时，$F=40.8$g；15°T 时，$F=21.7$g；30°T 时，$F=5.1$g；45°T 时，$F=0$g。

12.7.2　力—速度关系

最初将眼肌的粘性特性设计为非线性，是由于 Fenn 和 Marsh 的早期实验中没有发现预期的负载与最大速度之间存在的线性关系（见例 12.13）。正如 Fenn 和 Marsh 当时所指出的，"如果线性粘弹性系统能够准确描述眼肌，那么，其力—速度曲线应该是线性的，并且，力的损耗始终与速度成正比，而曲线的斜率就是粘性系数。"其实，1974 年 Close 和 Luff 用眼直肌重复进行了同样的实验，并得到类似的结果。

前面 12.6.4 节所述的检验眼肌粘弹性模型的经典力—速度实验具有两个假设。第一个假设是，眼肌等张收缩期间负载产生的惯性力可以忽略不计；第二个假设是，如果杠杆比不足以将惯性力减少到可以忽略的程度，那么就在最大速度时刻进行测量，此时加速度为 0，因此惯性力等于 0。如果这两个假设成立，那么，当负载重力变化时，所测得的肌张力实验数据就不会受到惯性力的影响。但是，实际上，这两个假设都不正确。第一个假设是错误的，因为惯性力不可能小到可以忽略的程度，必须予以考虑。第二个假设的错误是，在非 0 惯性质量的作用下，最大速度与达到此速度之前肌肉受到的作用力相关。本节将在考虑惯性力的情况下，重新仔细研究眼肌的力—速度关系。

线性眼肌模型的动态特性由力—速度曲线来描述，这种曲线可以用图 12.24 所示杠杆系统完成的等张实验数据求得。对于刚性杠杆，位移 x_1 和 x_3 直接与角度 θ_1 成正比，并且 x_1 与 x_3 相互之间也成正比，即

$$\theta_1=\frac{x_1}{d_1}=\frac{x_3}{d_3} \tag{12.50}$$

杠杆所受到的力矩为

$$Mgd_3+Md_3^2\ddot{\theta}_1=d_1K_{se}(x_2-x_1)+d_1B_2(\dot{x}_2-\dot{x}_1) \tag{12.51}$$

眼肌模型内部节点 2 上的作用力之和为

$$F=K_{lt}x_2+B_1\dot{x}_2+B_2(\dot{x}_2-\dot{x}_1)+K_{se}(x_2-x_1) \tag{12.52}$$

将式（12.50）代入式（12.51），消去 θ_1，可得

$$Mg\frac{d_3}{d_1}+M\left(\frac{d_3}{d_1}\right)^2\ddot{x}_1=K_{se}(x_2-x_1)+B_2(\dot{x}_2-\dot{x}_1) \tag{12.53}$$

理论上，要计算杠杆系统的力—速度曲线，就要先求解 $x_1(t)$；然后，用 $x_1(t)$ 求 $\dot{x}_1(t)$和$\ddot{x}_1(t)$；最后，根据 Close 和 Luff 的实验条件，求速度 $V_{max}=\dot{x}_1(T)$，此处时间 T 是肌肉收缩开始到停止所花的时间。其实，对于各个不同的数据点，该速度并不一定是最大速度，为了表示方便，这里就用符号 V_{max}表示力—速度曲线中的各个速度值。注意，此速度的定义与 Fenn 和 Marsh 的速度定义不同，Fenn 和 Marsh 将最大速度定义为 $V_{max}=\dot{x}_1(T)$，其中时间 T 由$\ddot{x}_1(T)=0$ 求得。

应该注意，这是一个三阶系统，$x_1(t)$ 的求解并不简单，涉及指数近似法。Enderle 等人（1988）发表的论文介绍了由四阶模型求 V_{max}指数近似解的方法，可供参考。还有一种权宜之计就是，直接用数值方法仿真求解 $x_1(t)$，然后用于计算随负载变化的 V_{max}函数。

仿真等张实验时，先从前述各个长度—张力曲线和眼直肌数据中计算出各个弹性元件的系数，各个粘性元件的参数则取自 Enderle 等人 1991 年发表的数值，即 $B_1=2.0$Ns/m 和 $B_2=0.5$Ns/m。其中，B_1由等张时间过程的时间常数求得；B_2则通过力—速度的仿真曲线与实验曲线的拟合求得，所取的 B_2值使两者拟合误差最小。

如图 12.31 所示，三角形表示利用式（12.53）求得的力—速度数据点，实线是经验拟合曲线。显然，拟合曲线与仿真数据匹配很好，并且，线性眼肌模型的力—速度曲线并不是直线。

式（12.51）和式（12.52）所描述的眼肌杠杆实验模型是一个三阶线性系统，具有 3 个极点。其特征值（即极点）的性质由参数值决定，可以是 3 个都是实数极点，也可以是一个实数极点加上一对复数共轭极点。实数极点是系统的主特征值。由灵敏度分析可知，粘性元件 B_1的数值对系统的主特征值（即时间常数）影响最大，而粘性元件 B_2则对主特征值没什么影响。因此，只要估计粘性元件 B_1的数值，使杠杆系统模型的主时间常数与等张实验数据的时间常数相匹配就可以了。当 $B_1>B_2$时，主时间常数约为 B_1/K_{lt}。对于眼直肌数据，等张实验的持续时间约为 100ms。取 $B_1=2.0$Ns/m 时，所得到的仿真等张响应曲线也具有约 100ms 的持续时间。对于骨

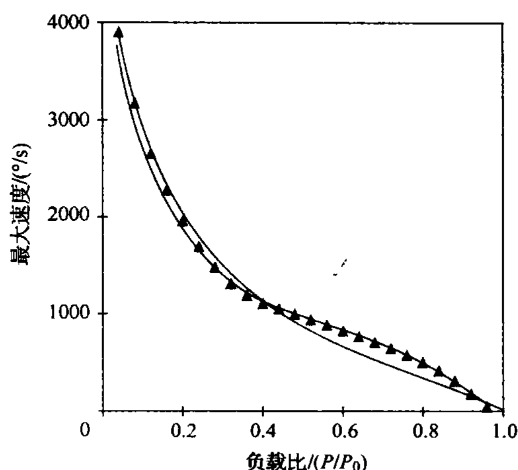


图 12.31　由线性眼肌模型仿真求得的力—速度曲线（摘自 Enderle、Engelken 和 Stiles，1991）
注：三角形表示仿真计算得到的数据点，实线表示 Close 和 Luff（1974）所述的经验拟合力—速度数据。

骼肌实验数据，等张实验的持续时间约为 400ms，取 $B_1=6.0$Ns/m 时，所产生的仿真等张响应曲线也具有相似的持续时间。我们知道快肌和慢肌具有不同形状的力—速度曲线，并且快肌的力—速度曲线的曲率较小。有趣的是，这里得到的两个 B_1 参数值分别给出了对应于快肌（即眼直肌）和慢肌（即骨骼肌）的两种不同力—速度曲线。

粘性元件 B_2 的参数值通过力—速度的仿真曲线与实验曲线的拟合求得，测试各种 B_2 的值使两者拟合误差达到最小。当 B_2 的值从 0.5Ns/m 逐渐减小时，力—速度曲线的形状越来越接近线性函数。而且，当 B_2 的值降到小于 0.3Ns/m 左右时，等张实验的仿真结果会出现强烈振荡，而实验数据中并不存在这种情况。因此，粘性元件 B_2 是肌模型的必要组成部分。如果没有它，力—速度曲线就接近线性，那么，等张模型的时间过程就与实验数据无法匹配。

通过改变杠杆实验眼肌模型的参数值，可以改变系统的特征值。例如，设 $M=0.5$kg，使用上述 B_1 和 B_2 参数值，可以求得系统的特征值为：一个实数极点 -30.71 和一对复数共轭极点 $-283.9\pm j221.2$。如果 B_2 的值增大，将会得到 3 个实数特征值；如果 B_2 的值减小，则还是一个实数极点和一对复数共轭极点。改变 B_1 的值并不改变特征值的组成，但却会显著改变主特征值的大小。例如，$B_1=0.1$ 时，主特征值为 -292；$B_1=6$ 时，主特征值就变为 -10。

12.8　线性同胚扫视眼动模型

根据生理现象，Bahill 等人于 1980 年提出了一个仿真水平眼动系统的四阶线性模型，此模型的仿真结果与水平眼动实验数据的匹配相当好，消除了 Westheimer 和

Robinson两种模型与实验数据之间在速度和加速度上的差异。为了便于叙述，下面用Enderle等人1984年发表的修正模型。

图12.32显示了水平眼动系统的力学组成结构，包括内、外直肌和眼球。主动肌用一个并联组合与一个弹性元件K_{se}的串联来模拟，该并联组合由激活态张力发生器F_{ag}、粘性元件B_{ag}和弹性元件K_{lt}组成。拮抗肌也用一个并联组合与一个弹性元件K_{se}的串联来模拟，其并联组合由激活态张力发生器F_{ant}、粘性元件B_{ant}和弹性元素K_{lt}组成。眼球模拟成转动惯量为J_p的球体，并与粘性元件B_p和弹性元件K_p相连。为了便于分析，各个眼肌的被动弹性特性都包含在K_p中。假设该动眼系统的每个元件都是理想且线性的。

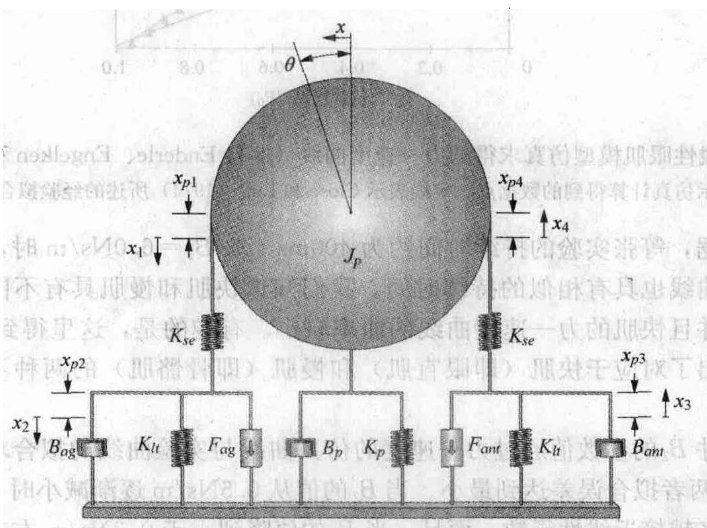


图12.32 动眼系统的力学结构示意图

注：图中所示各个眼肌处于偏离平衡位置的第一眼位，即眼睛直视前方，这与生理状态一致。在第一眼位时，眼直肌的平均长度约为40mm，而在平衡位置时约为37mm。θ为眼球偏离第一眼位的角度，变量x是弧线的长度。眼球处于第一眼位时，θ和x均为0。变量x_1～x_4分别是两块眼肌的各个刚性元件偏离平衡位置的位移。参数x_{p1}～x_{p4}分别是第一眼位下变量x_1～x_4偏离平衡位置的位移数值。第一眼位下眼肌偏离平衡位置的总长度变化是$x_{p1}+x_{p2}$，也就是$x_{p3}+x_{p4}$，约为3mm。假设内、外两块眼直肌相同，就有$x_{p1}=x_{p4}$，$x_{p2}=x_{p3}$。眼球半径为r。

图12.32所示眼肌模型的生理依据是前面12.6节所述的实验数据，力—速度曲线和弹性曲线都用线性拟合来近似。模型中眼球的转动惯量、弹性系数和粘性系数取自Robinson等人的斜视手术研究结果。眼球的被动弹性通常由其他4块动眼肌和视神经的作用共同产生。粘性作用则由眼球与眼眶的摩擦产生。转动惯量由眼球产生。

在图12.32所示模型中，分别求节点2和节点3的作用力之和，以及眼球所受力矩之和，再求工作点处的拉普拉斯变换，可以推导出该线性同胚模型的如下方程：

$$\delta(K_{se}(K_{st}(F_{ag}-F_{ant})+B_{ant}\dot{F}_{ag}-B_{ag}\dot{F}_{ant}))=\ddddot{\theta}+C_3\dddot{\theta}+C_2\ddot{\theta}+C_1\dot{\theta}+C_0\theta \quad (12.54)$$

其中

$$K_{st}=K_{se}+K_{lt},\ J=\frac{57.296J_p}{r^2},\ B=\frac{57.296B_p}{r^2},\ K=\frac{57.296K_p}{r^2},\ \delta=\frac{57.296}{rJB_{ant}B_{ag}},$$

$$C_3=\frac{JK_{st}(B_{ag}+B_{ant})+BB_{ant}B_{ag}}{JB_{ant}B_{ag}}$$

$$C_2=\frac{JK_{st}^2+BK_{st}(B_{ag}+B_{ant})+B_{ant}B_{ag}(K+2K_{se})}{JB_{ant}B_{ag}}$$

$$C_1=\frac{BK_{st}^2+(B_{ag}+B_{ant})(KK_{st}+2K_{se}K_{st}-K_{se}^2)}{JB_{ant}B_{ag}}$$

$$C_0=\frac{KK_{st}^2+2K_{se}K_{st}K_{lt}}{JB_{ant}B_{ag}}$$

设主动肌和拮抗肌的激活态张力与图 12.15 所示相同，并且没有潜伏期，则这两个张力的曲线就是如下低通滤波之后的波形：

$$\dot{F}_{ag}=\frac{N_{ag}-F_{ag}}{\tau_{ag}},\ \dot{F}_{ant}=\frac{N_{ant}-F_{ant}}{\tau_{ant}} \quad (12.55)$$

式中　N_{ag} 和 N_{ant}——两块眼肌的神经控制输入信号，即脉冲一阶跃信号；

τ_{ag} 和 τ_{ant}——时间常数，它们的计算公式是

$$\tau_{ag}=\tau_{ac}(u(t)-u(t-t_1))+\tau_{de}u(t-t_1) \quad (12.56)$$

$$\tau_{ant}=\tau_{de}(u(t)-u(t-t_1))+\tau_{ac}u(t-t_1) \quad (12.57)$$

根据实验研究结果，动眼系统的一组参数估计值为：$K_{se}=125\text{N/m}$，$K_{lt}=32\text{N/m}$，$K=66.4\text{N/m}$，$B=3.1\text{Ns/m}$，$J=2.2\times10^{-3}\text{Ns}^2/\text{m}$，$B_{ag}=3.4\text{Ns/m}$，$B_{ant}=1.2\text{Ns/m}$，$\tau_{ac}=0.009\text{s}$，$\tau_{de}=0.0054\text{s}$，$\delta=5.80288\times10^5$，且稳态激活态张力为

$$F_{ag}=\begin{cases}0.14+0.0185\theta & \text{N}\quad \theta<14.23°\text{时}\\ 0.0283\theta & \text{N}\quad \theta\geq14.23°\text{时}\end{cases} \quad (12.58)$$

$$F_{ant}=\begin{cases}0.14-0.009800\theta & \text{N}\quad \theta<14.23°\text{时}\\ 0 & \text{N}\quad \theta\geq14.23°\text{时}\end{cases} \quad (12.59)$$

由于每次扫视的变动很大，动态激活态张力一般要从连续的扫视过程中进行估算。估计激活态张力的一种方法就是使用系统辨识技术，它是频域上实现的一种共轭梯度搜索算法，本章后面将会介绍。每次扫视的主动肌脉冲幅值的变化很大，即使扫视幅度一样，输入信号中的脉冲幅值也有很大的差别。并且，主动肌脉冲的持续时间与脉冲幅值密切相关。对于同样的扫视幅度，脉冲幅值增加，脉冲持续时间就减小。对于图 12.32 所示模型，脉冲幅值的合理取值范围约为 0.6N～1.4N。脉冲幅值越大，扫视峰值也越大。

例 12.15　请根据图 12.32 所示模型推导方程式（12.54）。

解： 由图 12.32 可知，变量 x_1、x_4、θ 和 x 之间存在如下关系：

$$x_1=x-x_{p1}$$

$$x_4=x+x_{p4}$$

$$\theta=57.296x/r$$

假设$\dot{x}_2>\dot{x}_1\geqslant\dot{x}_4>\dot{x}_3$，则节点 2 和节点 3 上的作用力之和，以及眼球所受到的力矩之和分别为

$$F_{ag}=B_{ag}\dot{x}_2+K_{se}(x_2-x_1)+K_{lt}x_2$$

$$rK_{se}(x_2-x_1)-rK_{se}(x_4-x_3)=J_p\ddot{\theta}+B_p\dot{\theta}+K_p\theta$$

$$K_{se}(x_4-x_3)=F_{ant}+K_{lt}x_3+B_{ant}\dot{x}_3$$

求每个方程的拉普拉斯变换，然后，消去 x 以外的其他变量，就可以将此 3 个方程简化为一个方程。不过，拉普拉斯变换需要初始条件。还有一种方法是将方程简化为只包含相对于初始眼球位置的变化量，即引入

$$\hat{x}=x-x(0)$$

$$\hat{\theta}=\theta-\theta(0)$$

$$\hat{x}_1=x_1-x_1(0)$$

$$\hat{x}_2=x_2-x_2(0)$$

$$\hat{x}_3=x_3-x_3(0)$$

$$\hat{x}_4=x_4-x_4(0)$$

$$\hat{F}_{ag}=F_{ag}-F_{ag}(0)$$

$$\hat{F}_{ant}=F_{ant}-F_{ant}(0)$$

再将以上节点 2、3 的作用力方程和眼球力矩的方程用新变量重新写成

$$\hat{F}_{ag}=K_{st}\hat{x}_2+B_{ag}\dot{\hat{x}}_2-K_{se}\hat{x}$$

$$\hat{F}_{ant}=K_{se}\hat{x}-K_{st}\hat{x}_3-B_{ant}\dot{\hat{x}}_3$$

$$K_{se}(\hat{x}_2+\hat{x}_3-2\hat{x})=J\ddot{\hat{x}}+B\dot{\hat{x}}+K\hat{x}$$

其中

$$K_{st}=K_{se}+K_{lt}$$

$$J=\frac{57.296J_p}{r^2}$$

$$B=\frac{57.296B_p}{r^2}$$

$$K=\frac{57.296K_p}{r_2}$$

注意，由于变量 $\hat{x}_1$和 $\hat{x}_4$都等于 x，且 $\hat{\theta}=\dfrac{57.296\hat{x}}{r}$，这几个变量已从以上方程中消去。

求以上 3 个方程的拉普拉斯变换，可得

$$L\{\hat{F}_{ag}\}=(K_{st}+sB_{ag})\hat{x}_2-K_{se}\hat{x}$$

$$L\{\hat{F}_{ant}\}=K_{se}\hat{x}-(K_{se}+sB_{ag})\hat{x}_3$$

$$K_{se}(\hat{x}_2+\hat{x}_3-2\hat{x})=(Js^2+Bs+K)\hat{x}$$

从前 2 个方程分别求得 $\hat{x}_2$ 和 $\hat{x}_3$，即

$$\hat{x}_2=\frac{L\{\hat{F}_{ag}\}+K_{se}\hat{x}}{K_{st}+sB_{ag}}$$

$$\hat{x}_3=\frac{K_{se}\hat{x}-L\{\hat{F}_{ant}\}}{K_{st}+sB_{ant}}$$

代入第 3 个方程，得到

$$K_{se}\left[\frac{L\{\hat{F}_{ag}\}+K_{se}\hat{x}}{K_{st}+sB_{ag}}+\frac{K_{se}\hat{x}-L\{\hat{F}_{ant}\}}{K_{st}+sB_{ant}}-2\hat{x}\right]=(s^2J+sB+K)\hat{x}$$

将此方程两边同乘以 $(sB_{ant}+K_{st})\times(sB_{ag}+K_{st})$，简化为

$$K_{se}(sB_{ant}+K_{st})L\{\hat{F}_{ag}\}-K_{se}(sB_{ag}+K_{st})L\{\hat{F}_{ant}\}=(P_4s^4+P_3s^3+P_2s^2+P_1s+P_0)\hat{x}$$

其中

$$P_4=JB_{ant}B_{ag}$$
$$P_3=JK_{st}(B_{ag}+B_{ant})+BB_{ant}B_{ag}$$
$$P_2=JK_{st}^2+BK_{st}(B_{ag}+B_{ant})+B_{ant}B_{ag}(K+2K_{se})$$
$$P_1=BK_{st}^2+(B_{ag}+B_{ant})(KK_{st}+2K_{se}K_{st}-K_{se}^2)$$
$$P_0=KK_{st}^2+2K_{se}K_{st}K_{lt}$$

将其重新变换到时域，可得

$$K_{se}(K_{st}(\hat{F}_{ag}-\hat{F}_{ant})+B_{ant}\dot{\hat{F}}_{ag}-B_{ag}\dot{\hat{F}}_{ant})=P_4\ddddot{\hat{x}}+P_3\dddot{\hat{x}}+P_2\ddot{\hat{x}}+P_1\dot{\hat{x}}+P_0\hat{x}$$

注意，由于 $\dot{F}_{ag}=\dot{\hat{F}}_{ag}$，$\dot{F}_{ant}=\dot{\hat{F}}_{ant}$，且 $\hat{x}$ 的导数等于 x 的导数，该方程可以写为

$$K_{se}(K_{st}(\hat{F}_{ag}-\hat{F}_{ant})+B_{ant}\dot{F}_{ag}-B_{ag}\dot{F}_{ant})=P_4\ddddot{x}+P_3\dddot{x}+P_2\ddot{x}+P_1\dot{x}+P_0\hat{x}$$

一般，$Fag(0)$ 不一定等于 $F_{ant}(0)$。由于 $\hat{x}=x-x(0)$，且 $\hat{F}_{ag}-\hat{F}_{ant}=F_{ag}-F_{ag}(0)-F_{ant}+F_{ant}(0)$，两个节点方程的稳态值之差为

$$F_{ag}(0)-F_{ant}(0)=K_{st}(x_2(0)+x_3(0))-K_{se}(x_1(0)+x_4(0))$$

由于 $x_1(0)=x(0)-x_{p1}$ 且 $x_4(0)=x(0)+x_{p1}$，此式可以写为

$$F_{ag}(0)-F_{ant}(0)=K_{st}(x_2(0)+x_3(0))-2K_{se}x(0)$$

在 0^- 时刻，节点 1 方程为

$$K_{se}(x_2(0)+x_3(0)-2x(0))=Kx(0)$$

即

$$x_2(0)+x_3(0)=\left(\frac{K}{K_{se}}+2\right)x(0)$$

由此可得

$$F_{ag}(0)-F_{ant}(0)=\left(K_{st}\left(\frac{K}{K_{se}}+2\right)-2K_{se}\right)x(0)$$

因此

$$K_{se}(K_{st}(F_{ag}-F_{ant})-K_{st}(F_{ag}(0)-F_{ant}(0))+B_{ant}\dot{F}_{ag}-B_{ag}\dot{F}_{ant})$$

$$=P_4\ddddot{x}+P_3\dddot{x}+P_2\ddot{x}+P_1\dot{x}+P_0(x-x(0))$$

将以上 $F_{ag}(0)-F_{ant}(0)$ 表达式两边同乘以 $K_{se}K_{st}$，得到

$$K_{se}K_{st}(F_{ag}(0)-F_{ant}(0))=K_{se}K_{st}\left(K_{st}\left(\frac{K}{K_{se}}+2\right)-2K_{se}\right)x(0)$$

简化此式右边，可得

$$K_{se}K_{st}(F_{ag}(0)-F_{ant}(0))=(K_{st}^2K+2K_{se}K_{st}^2-2K_{se}^2K_{st})x(0)$$
$$=(K_{st}^2K+2K_{se}K_{st}K_{lt})x(0)=P_0x(0)$$

将上式代入系统的微分方程，得到

$$K_{se}(K_{st}(F_{ag}-F_{ant})+B_{ant}\dot{F}_{ag}-B_{ag}\dot{F}_{ant})=P_4\ddddot{x}+P_3\dddot{x}+P_2\ddot{x}+P_1\dot{x}+P_0x$$

再将 $x=\theta r/57.296$ 代入上式，就得到方程（12.54）。 ■

12.9 更真实的线性同胚扫视眼动模型

12.8 节介绍的动眼系统线性模型是 Hsu 等人的非线性动眼系统模型的线性化结果，也就是将力一速度曲线和弹性曲线进行了线性化处理。如果直接利用 12.7 节所述的眼肌线性模型，就不需要线性化处理，便可以推导出更真实的线性同胚扫视眼动模型。

12.7 节所述的线性眼肌模型包含了眼直肌静态特性和动态特性，并且没有任何非线性元件。如前所述，该模型利用线性粘性元件模拟出一种非线性力一速度关系曲线，此曲线与眼肌的实验数据相符。并且，模型的长度一张力特性也与工作范围内的眼肌实验数据非常吻合。另外，线性眼肌模型还有其他优势，例如，如果平衡位点设为 $x_e=-19.3°$，而不是 15°，那么，模型中就不需要用被动弹性元件；并且，肌肉的粘性系数是一个常数，与神经控制刺激信号的强弱无关。

图 12.33 显示了新的水平眼动系统的力学组成结构，包括内、外眼直肌和眼球。主动肌用两个并联组合的串联来模拟，其中，一个并联组合由粘性元件 B_2 和串联弹性元件 K_{se} 组成，另一个并联组合由激活态张力发生器 F_{ag}、粘性元件 B_1 和长度一张力弹性元件 K_{lt} 组成。为了简化，设主动肌粘性系数与拮抗肌粘性系数相等。拮抗肌的模型与主动肌相同，只是将其激活态张力改为 F_{ant} 即可。假设该动眼系统的每个元件都是理想且线性的。

眼球用转动惯量为 J_p 的球体模拟，并与两个串联连接的粘弹性元件相连。该眼球模型根据 Robinson 于 1981 年发表的实验结果建立，介绍如下。在 12.8 节所述的眼动系统模型中，被动弹性元件 K_{pe} 与眼眶组织的被动弹性特性结合在一起。而 12.7 节所述的线性眼肌模型不包含弹性元件 K_{pe}。因此，这里要根据 Robinson 的实验结果，重新设定眼窝组织的被动弹性特性。Robinson 在 1981 年指出，“断开水平直肌后，如果将人眼移动一个距离，然后突然放开，眼球会以 0.02s 左右的时间常数迅速恢复约 61%的位移，而后，再以 1s 左右的时间常数缓慢回复到原位。”该实验现象表明，该系统中至少存在两个粘弹性元件。因此，这里用两对粘弹性元件的串联组

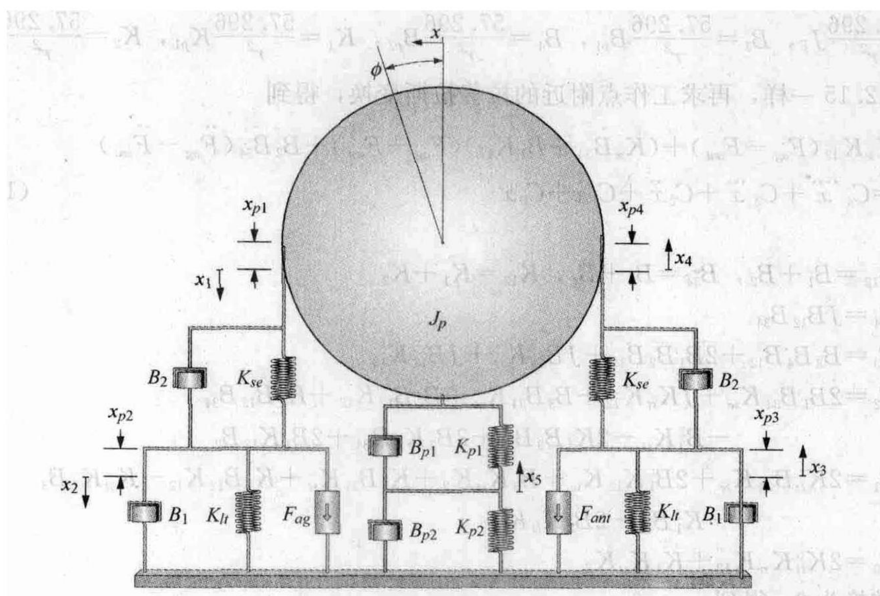


图 12.33　新的水平眼动系统力学结构示意图

注：图中所示各个眼肌处于偏离平衡位置的第一眼位，即眼睛直视前方，与生理现象一致。眼直肌在第一眼位的平均长度约为 40mm，而在平衡位置时约为 37mm。θ 为眼球偏离第一眼位的角度，变量 x 是弧线的长度。眼球处于第一眼位时，θ 和 x 均为 0。变量 $x_1 \sim x_4$ 分别表示两块眼肌的各个刚性元件偏离平衡位置的位移，θ_5 是眼眶组织的被动旋转位移。参数值 $x_{p1} \sim x_{p4}$ 分别是第一眼位下变量 $x_1 \sim x_4$ 偏离平衡位置的位移值。第一眼位下眼肌偏离平衡位置的总长度变化是 $x_{p1}+x_{p2}$，也就是 $x_{p3}+x_{p4}$，约为 3mm。假设内、外两块眼直肌相同，就有 $x_{p1}=x_{p4}$，$x_{p2}=x_{p3}$。眼球半径为 r。

合来代替前一节所述模型中的一对粘弹性元件，并与眼球相连。这两对粘弹性元件分别由 B_{p1} 与 K_{p1} 并联以及 B_{p2} 与 K_{p2} 并联构成。如 Robinson 所提出的，眼眶的总弹性系数为 12.8×10^{-7}g/(°)（按比例计算）。于是，根据上述时间常数，眼眶各个粘弹性元件的参数分别为 $K_{p1}=1.28\times10^{-6}$g/(°)，$K_{p2}=1.98\times10^{-6}$g/(°)，$B_{p1}=2.56\times10^{-8}$g/(°)，$B_{p2}=1.98\times10^{-6}$g/(°)。为了方便建模，用变量 θ_5 表示这两对粘弹性元件偏离平衡位置的变化量。θ 和 θ_5 这两个变量都不出现在模型方程中，它们将被 $\theta=57.296x/r$ 和 $\theta_5=57.296x_5/r$ 代替。

在图 12.33 所示模型中，分别求节点 2 和节点 3 的作用力之和、眼球所受力矩之和以及节点 5 的作用力之和，可以得到如下 4 个方程组成的方程组：

$$
\begin{aligned}
F_{ag} &= K_{lt}x_2 + B_1\dot{x} + K_{se}(x_2 - x_1) + B_2(\dot{x}_2 - \dot{x}_1) \\
B_2(\dot{x}_4 - \dot{x}_3) + K_{se}(x_4 - x_3) &= F_{ant} + K_{lt}x_3 + B_1\dot{x}_3 \\
B_2(\dot{x}_2 + \dot{x}_3 - \dot{x}_1 - \dot{x}_4) + K_{se}(x_2 + x_3 - x_1 - x_4) &= J\ddot{x} + B_3(\dot{x} - \dot{x}_5) + K_1(x - x_5) \\
K_1(x - x_5) + B_3(\dot{x} - \dot{x}_5) &= B_4\dot{x}_5 + K_2x_5
\end{aligned}
\tag{12.60}
$$

其中

$J=\frac{57.296}{r^2}J_p$，$B_3=\frac{57.296}{r^2}B_{p1}$，$B_4=\frac{57.296}{r^2}B_{p2}$，$K_1=\frac{57.296}{r^2}K_{p1}$，$K_2=\frac{57.296}{r^2}K_{p2}$

与例 12.15 一样，再求工作点附近的拉普拉斯变换，得到

$$K_{se}K_{12}(F_{ag}-F_{ant})+(K_{se}B_{34}+B_2K_{12})(\dot{F}_{ag}-\dot{F}_{ant})+B_2B_{34}(\ddot{F}_{ag}-\ddot{F}_{ant})$$
$$=C_4\ddddot{x}+C_3\dddot{x}+C_2\ddot{x}+C_1\dot{x}+C_0x \tag{12.61}$$

其中

$$B_{12}=B_1+B_2,\ B_{34}=B_3+B_4,\ K_{12}=K_1+K_2$$
$$C_4=JB_{12}B_{34}$$
$$C_3=B_3B_4B_{12}+2B_1B_2B_{34}+JB_{34}K_{st}+JB_{12}K_{12}$$
$$C_2=2B_1B_{34}K_{se}+JK_{st}K_{12}+B_3B_{34}K_{st}+B_3B_{12}K_{12}+K_1B_{12}B_{34}$$
$$-B_3^2K_{st}-2K_1B_3B_{12}+2B_2K_{lt}B_{34}+2B_1K_{12}B_2$$
$$C_1=2K_{lt}B_{34}K_{se}+2B_1K_{12}K_{se}+B_3K_{st}K_2+K_1B_{34}K_{st}+K_1B_{12}K_{12}-K_{st}K_1B_3$$
$$-K_1^2B_{12}+2B_2K_{lt}K_{12}$$
$$C_0=2K_{lt}K_{se}K_{12}+K_1K_{st}K_2$$

将 x 转换为 θ，得到

$$\delta(K_{se}K_{12}(F_{ag}-F_{ant})+(K_{se}B_{34}+B_2K_{12})(\dot{F}_{ag}-\dot{F}_{ant})+B_2B_{34}(\ddot{F}_{ag}-\ddot{F}_{ant}))$$
$$=\ddddot{\theta}+P_3\dddot{\theta}+P_2\ddot{\theta}+P_1\dot{\theta}+P_0\theta \tag{12.62}$$

其中

$$\delta=\frac{57.296}{rJB_{12}B_{34}},\ P_3=\frac{C_3}{C_4},\ P_2=\frac{C_2}{C_4},\ P_1=\frac{C_1}{C_4},\ P_0=\frac{C_0}{C_4}$$

根据 12.6 节所述的改进眼肌模型及其长度－张力数据，可以计算出稳态时的激活态张力为

$$F=\begin{cases}0.4+0.01750\theta & N \quad \theta\geqslant 0°\text{时}\\ 0.4+0.01250\theta & N \quad \theta<0°\text{时}\end{cases} \tag{12.63}$$

设主动肌和拮抗肌的激活态张力与图 12.15 所示相同，并且没有潜伏期，则这两个张力的曲线就是如下低通滤波之后的波形：

$$\dot{F}_{ag}=\frac{N_{ag}-F_{ag}}{\tau_{ag}},\ \dot{F}_{ant}=\frac{N_{ant}-F_{ant}}{\tau_{ant}} \tag{12.64}$$

式中 N_{ag} 和 N_{ant}——主动肌和拮抗肌神经控制输入信号，即脉冲一阶跃波形；

τ_{ag} 和 τ_{ant}——时间常数，其计算公式为

$$\tau_{ag}=\tau_{ac}(u(t)-u(t-t_1))+\tau_{de}u(t-t_1) \tag{12.65}$$
$$\tau_{ant}=\tau_{de}(u(t)-u(t-t_1))+\tau_{ac}u(t-t_1) \tag{12.66}$$

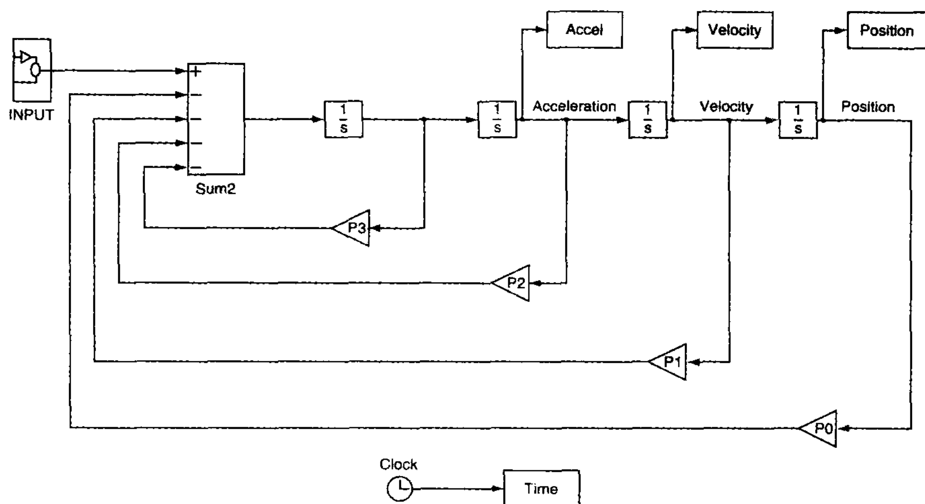
根据实验数据的分析结果，动眼系统的各个参数分别为：$K_{se}=125\text{N/m}$，$K_{lt}=60.7\text{N/m}$，$B_1=2.0\text{Ns/m}$，$B_2=0.5\text{Ns/m}$，$J=2.2\times10^{-3}\text{Ns}^2/\text{m}$，$B_3=0.538\text{Ns/m}$，$B_4=41.54\text{Ns/m}$，$K_1=26.9\text{N/m}$，$K_2=41.54\text{N/m}$，$r=0.0118\text{m}$。

该模型仿真所得到的扫视眼动各项特性，如位置、速度、加速度和各个主特性

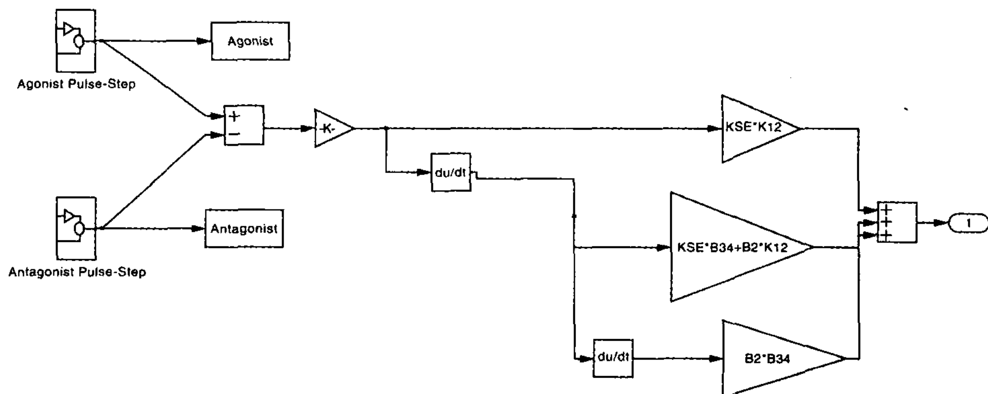
曲线，都与实验数据匹配得很好。与前一节结果相同，主动肌脉冲幅值与脉冲持续时间两者之间紧密相关。

例 12.16　请应用方程（12.62）的眼动系统模型及其后面所列的相关参数，以及方程（12.63）表示的稳态输入信号，仿真 10°扫视眼动过程。作出主动肌和拮抗肌的激活态张力、位置、速度和加速度随时间变化的曲线，并将仿真结果与图 12.11 所示的主特性曲线进行比较。

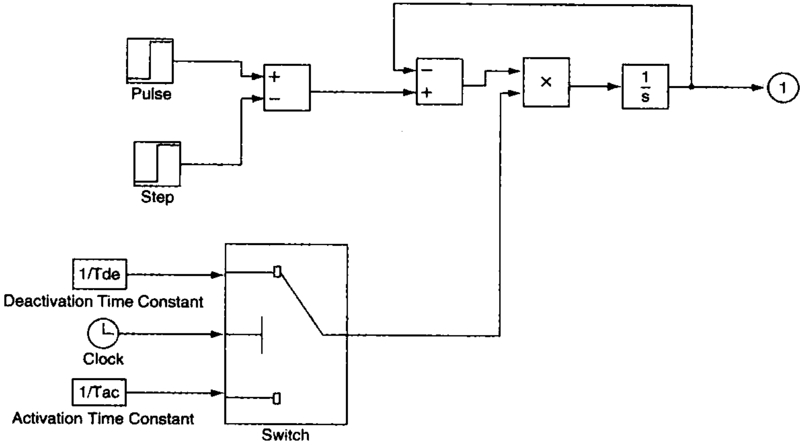
解：本题求解的关键是要选择一组参数值，包括 F_p、t_1、τ_{ac}、τ_{de}等，使模型的仿真结果可以与图 12.11 所示主特性曲线相匹配。10°扫视的仿真可以用很多不同的参数值来实现，其约束条件只有扫视持续时间约为 40～50ms，最大速度范围为500～600°/s。根据实际情况，再加上 150ms 潜伏期。下图所示是方程（12.62）的几个 SIMULINK 仿真模块图。



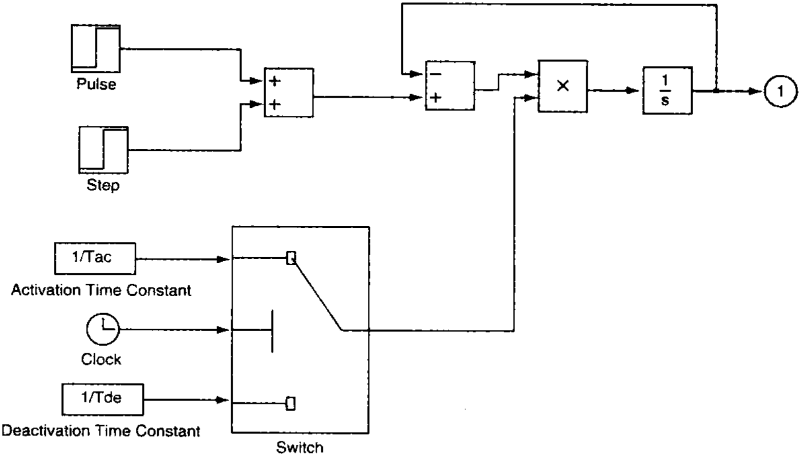
其中输入模块为



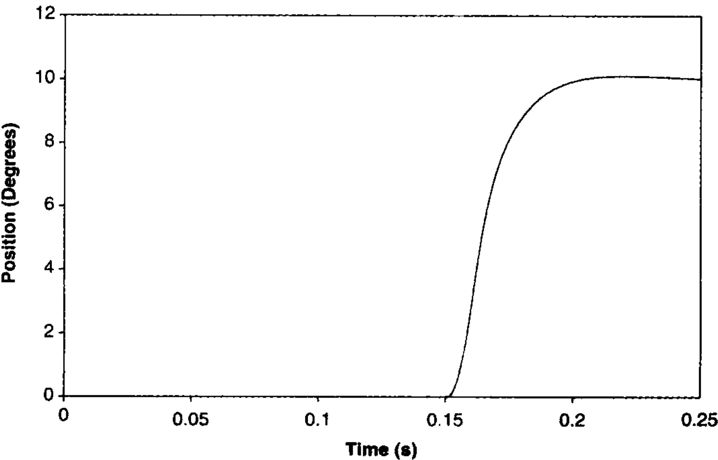
主动肌的脉冲一阶跃信号模块为



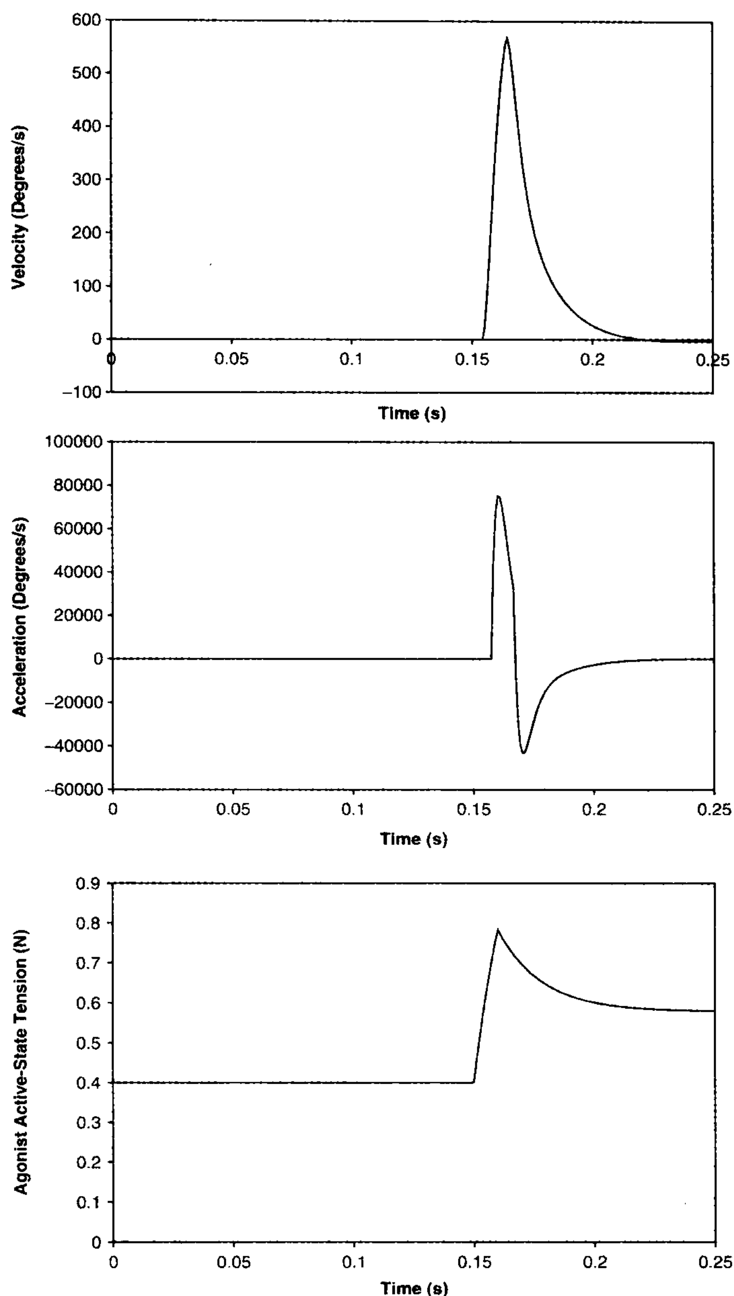
拮抗肌的脉冲—阶跃信号模块为



已知 $F_p=1.3\text{N}$，$t_1=0.10\text{s}$，$\tau_{ac}=0.018\text{s}$，$\tau_{de}=0.018\text{s}$，下面是仿真结果。

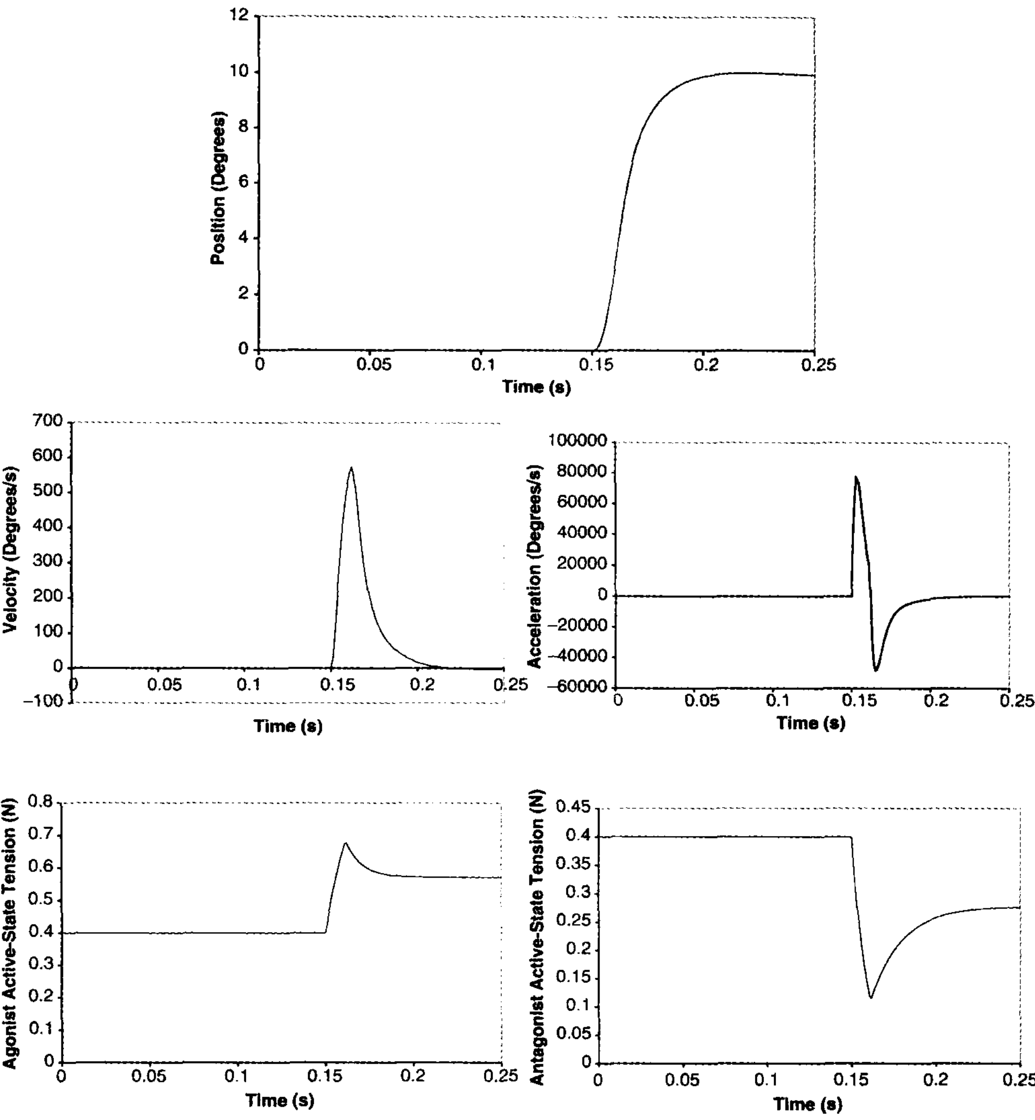


以下是该 10°扫视仿真的各个主特性曲线，其中最大速度为 568°/s，持续时间为 45ms。

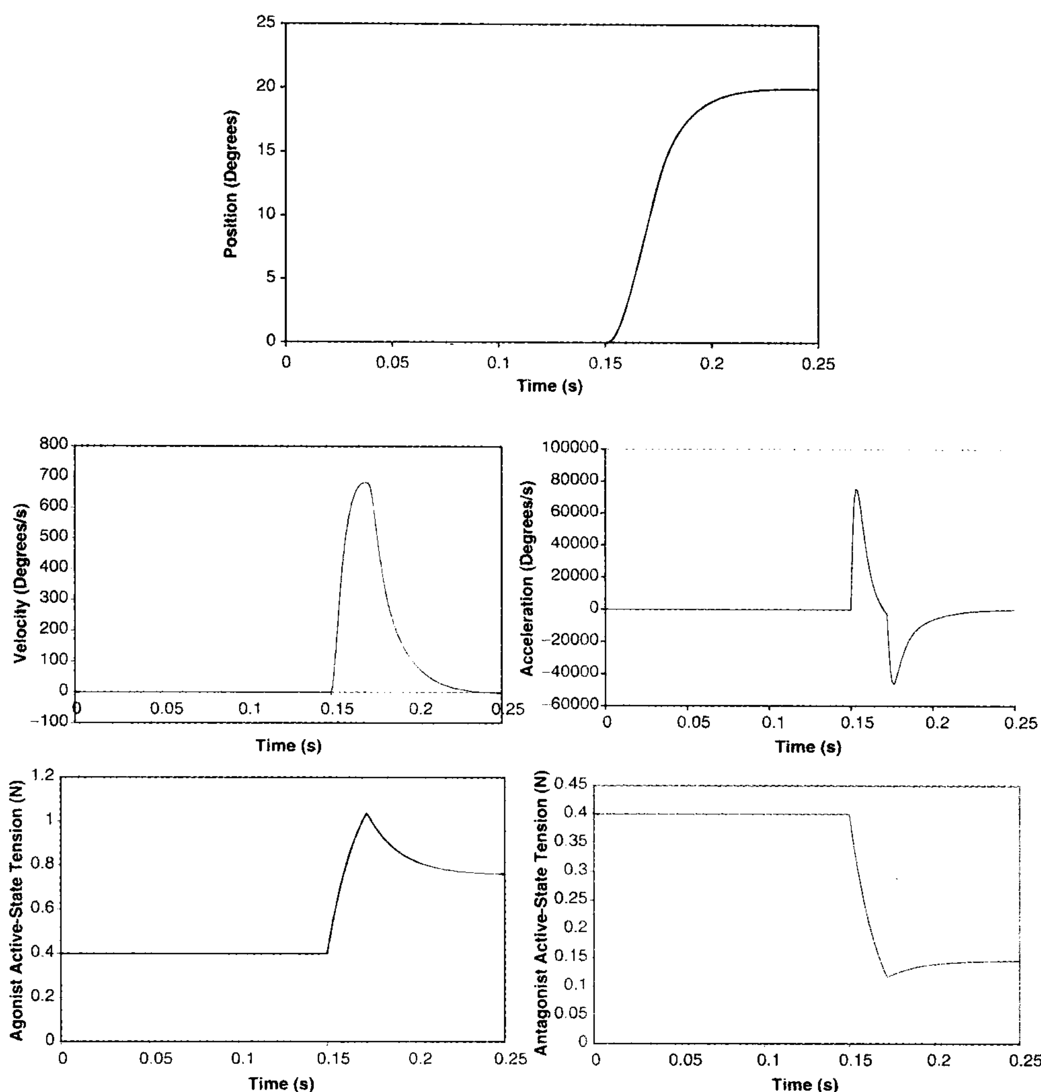


用其他参数值也可以仿真 10°扫视眼动过程。例如，可以将时间常数 τ_{de} 减小为 0.009s。这样，拮抗肌的激活态张力下降到 0 的速度就比以上仿真要快，导致总激活

态张力 $F_{ag}-F_{ant}$增大。因此，如果 τ_{ac}仍为 0.018s，且 $t_1=0.0115$s，那么，要使 10°扫视仍然具有合理的主特性曲线，F_p就必须减小到 1.0N。此时 10°扫视的仿真结果如下图所示。



如果要仿真较大幅度扫视活动的主特性曲线，主动肌和拮抗肌激活态张力的时间常数可以保持不变，与 10°扫视时的取值一样，也可以设定为扫视幅度的函数（参见 Bahill 在 1981 年提出的几种与扫视幅度相关的时间常数）。取 $F_p=1.3$N，两个时间常数仍然为 0.018s，并分别设 $t_1=0.0155$s 和 $t_1=0.0223$s，就可以仿真 15°和 20°扫视眼动的各个主特性曲线。下图显示了 20°扫视的仿真结果，其中最大速度为 682°/s，持续时间为 60ms。



总之，要维持同样的扫视幅度，如果 F_p 增加了，t_1 就要减小。另外，最大速度会随着 F_p 的增大而增加。如果两个时间常数之一或者两者都增加，那么，要保持扫视幅度不变，F_p 也要增加。■

12.10　系统辨识

在传统电气工程、机械工程和化学工程领域中，建模仿真主要是作为一种低成本的设计工具，用于有效地研究各种参数变化对于系统性能的影响。而生理系统的建模并不是为了设计一个系统，而是为了确定系统的各个参数及其组成结构。理论

上，如图 12.34 所示，如果已知生理系统的输入和输出，就可以获得系统内部的某些动态特性。但是，很多情况下，不是输入测不到，就是输出不可测，只能通过其他外围信号间接估计系统的输入或者输出，并且也不知道系统内部的任何信息。系统辨识就是通过建立系统模型，来估计系统各个参数的方法。本节将分别介绍频域和时域上系统辨识的基本原理。

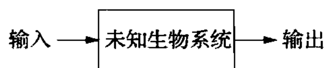


图 12.34 无反馈的典型生物系统框图

生物医学工程师可以测得各种各样信号，例如，人体产生的动作电位、脑电图（EEG）、心电图（ECG）、肌电图（EMG）、眼电图（EOG）等电信号，以及由压力传感器测得的非电信号，还有通过超声波、X 射线断层摄影、磁共振成像（MRI）等获得的各种图像信号。有了这些信号，就可以通过图 12.1 所示的建模流程图来建立系统模型，并估计模型的各个参数。在开始进行系统辨识之前，必须先了解系统输入和输出信号的特性，也就是，弄清楚电压范围、频率范围、信号是确定性的还是随机性的、是否是编码信号（如神经编码）等问题。生物体所产生的大部分信号都属于低频信号，并且都涉及一定的编码过程。例如，脑电图（EEG）的上限频率为 30 Hz，而眼睛运动的上限频率小于 100 Hz。本章前面介绍的眼球扫视系统则通过神经编码将爆发式神经脉冲发放的持续时间转换为眼球的扫视幅度。获得输入信号和输出信号之后，必须对这些信号进行处理。最基本的处理器就是放大器，如本书第 8 章所述，增益和频率响应是放大器的基本特性。注意，为了消除高频噪声的干扰，一般放大器都设计成低通滤波器（Low-Pass Filter，LPF）。有趣的是，大部分放大器都具有储能元件（即电容和电感），因此，必须等到放大器的暂态响应结束之后，才能采集到有用的信号。记住，滤波器的频域截止速度越快，放大器的暂态响应时间就越长。

本课程学习中，经常会遇到这样的问题，就是在给定系统传递函数和输入信号的条件下，求解系统的输出响应。这看起来虽然有点难，但是，实际上比这里将要介绍的确定生理系统参数的问题要容易得多。计算生理系统参数时，只知道模型的输入信号以及含噪声的输出信号，并且，很可能像扫视眼动系统那样，已知的并不是直接的输入信号，而是间接信号。理论上，生理系统建模所要达到的理想结果就是要获得下式表示的系统传递函数

$$H(s)=\frac{V_{\mathrm{o}}(s)}{V_{\mathrm{i}}(s)} \tag{12.67}$$

12.10.1 传统系统辨识方法

系统辨识最简单且最直接的方法就是正弦输入分析法。相应的测试装置通常包括：正弦激励器、测量传感器以及采集频率响应输出信号的记录仪。如本书第 9 章所述，现在有很多各式各样的传感器，可以将各种物理变量转换成电信号。但是，现成的正弦激励装置却很难找到，一般都需要研究人员自己动手设计。频率响应输出信号则用示波器就可以记录。图 12.35 显示了正弦输入分析法的原理。

应用正弦输入分析法辨识模型参数的实验很简单。只要取整个感兴趣频率范围

$V_i = A\cos(\omega_x t+\theta)$ → $h(t)$ → $V_o = B\cos(\omega_x t+\phi)$

图 12.35　系统正弦输入响应框图

注：V_i为输入信号，V_o为输出信号，$h(t)$为系统冲激响应。

内的一系列离散频率点上的单频正弦波信号作为输入，同时在系统输出端检测对应于每个正弦输入的输出正弦波的幅值和相位。下面介绍这种方法的基本原理。由图 12.35 可知，系统的传递函数为

$$H(\mathrm{j}\omega)=\frac{V_o(\mathrm{j}\omega)}{V_i(\mathrm{j}\omega)} \tag{12.68}$$

输入信号的傅里叶变换为

$$V_i(\mathrm{j}\omega)=F\{A\cos(\omega_x t+\theta)\}=A\int \mathrm{e}^{-\mathrm{j}\omega\lambda}\cos(\omega_x\lambda+\theta)\mathrm{d}\lambda=A\int \mathrm{e}^{-\mathrm{j}\omega\left(-\frac{\theta}{\omega_x}+\tau\right)}\cos(\omega_x\tau)\mathrm{d}\tau \tag{12.69}$$

将 $\lambda=\tau-\dfrac{\theta}{\omega_x}$ 代入，并提出不含 τ 的因子项，可得

$$V_i(\mathrm{j}\omega)=A\mathrm{e}^{\frac{\mathrm{j}\omega\theta}{\omega_x}}\int \mathrm{e}^{-\mathrm{j}\omega\tau}\cos\omega_x\tau d\tau=A\mathrm{e}^{\frac{\mathrm{j}\omega\theta}{\omega_x}}[\pi\delta(\omega-\omega_x)+\pi\delta(\omega+\omega_x)] \tag{12.70}$$

同理，输出信号的傅里叶变换为

$$V_o(\mathrm{j}\omega)=B\mathrm{e}^{\frac{\mathrm{j}\omega\phi}{\omega_x}}\ [\pi\delta\ (\omega-\omega_x)\ +\pi\delta\ (\omega+\omega_x)] \tag{12.71}$$

将以上输入信号和输出信号代入式（12.68)，可得

$$H(\mathrm{j}\omega)=\frac{V_o(\mathrm{j}\omega)}{V_i(\mathrm{j}\omega)}=\frac{B}{A}\frac{\mathrm{e}^{\frac{\mathrm{j}\omega\phi}{\omega_x}}}{\mathrm{e}^{\frac{\mathrm{j}\omega\theta}{\omega_x}}}=\frac{B}{A}\mathrm{e}^{\frac{\mathrm{j}\omega(\phi-\theta)}{\omega_x}} \tag{12.72}$$

稳态时，$\omega=\omega_x$，上式可以简化为

$$H(\mathrm{j}\omega)=\frac{B}{A}\mathrm{e}^{\mathrm{j}(\phi-\theta)} \tag{12.73}$$

此式中的 A、B、ϕ 和 θ 都是已知量，因此，$\omega=\omega_x$时的传递函数幅值和相位就可以求得。这样，在感兴趣的频率范围内改变 ω_x的值，就可以求得整个传递函数。

一般，传递函数 $G(s)\Big|_{s=\mathrm{j}\omega}=G\ (\mathrm{j}\omega)$ 由以下几项组成：

1）常数项 K。

2）原点处的 M 个极点，即零点，其形式为 $(\mathrm{j}\omega)^M$。

3）形式为 $\Pi_{p=1}^{P}(1+\mathrm{j}\omega\tau_p)$ 的 P 个极点，或者形式为 $\Pi_{z=1}^{Z}(1+\mathrm{j}\omega\tau_z)$ 的 Z 个零点。实际上，这些极点（或零点）都位于$-\dfrac{1}{\tau}$。

4）形式为 $\Pi_{r=1}^{R}\left(1+\left(\dfrac{2\zeta_r}{\omega_{n_r}}\right)\mathrm{j}\omega+\left(\dfrac{\mathrm{j}\omega}{\omega_{n_r}}\right)^2\right)$ 的 R 个复数极点，或者形式为 $\prod\limits_{s=1}^{S}\left(1+\left(\dfrac{2\zeta_s}{\omega_{n_s}}\right)\mathrm{j}\omega+\left(\dfrac{\mathrm{j}\omega}{\omega_{n_s}}\right)^2\right)$的 S 个零点。

5）纯时延 $\mathrm{e}^{-\mathrm{j}\omega T_d}$。

其中，M、P、Z、R、S 和 T_d 都是正整数。将这些项组合在一起就是如下的传递函数

$$G(j\omega)=\frac{K\times e^{-j\omega T_d}\times(\Pi_{z=1}^{Z}(1+j\omega\tau_z))\times\left(\Pi_{s=1}^{S}\left(1+\left(\frac{2\zeta_s}{\omega_{n_s}}\right)j\omega+\left(\frac{j\omega}{\omega_{n_s}}\right)^2\right)\right)}{(j\omega)^M\times(\Pi_{p=1}^{P}(1+j\omega\tau_p))\times\left(\Pi_{r=1}^{R}\left(1+\left(\frac{2\zeta_r}{\omega_{n_r}}\right)j\omega+\left(\frac{j\omega}{\omega_{n_r}}\right)^2\right)\right)} \tag{12.74}$$

此方程是系统建模的模板。根据传递函数测量数据的对数值和渐近线，可以计算出方程中的各个未知参数。一般，传递函数的对数增益（即输出与输入的幅值比）定义为（单位为 dB，对数是以 10 为基底）：

$$\begin{aligned}20\lg|G(j\omega)| &= 20\lg K+20\sum_{z=1}^{Z}\lg|1+j\omega\tau_z|+20\sum_{s=1}^{S}\lg\left|1+\left(\frac{2\zeta_s}{\omega_{n_s}}\right)j\omega+\left(\frac{j\omega}{\omega_{n_s}}\right)^2\right|\\&\quad-20\lg|(j\omega)^M|-20\sum_{p=1}^{P}\lg|1+j\omega\tau_p|-20\sum_{r=1}^{R}\lg\left|1+\left(\frac{2\zeta_r}{\omega_{n_r}}\right)j\omega+\left(\frac{j\omega}{\omega_{n_r}}\right)^2\right|\end{aligned} \tag{12.75}$$

相位定义为（单位为°）：

$$\begin{aligned}\phi(\omega) &=-\omega T_d+\sum_{z=1}^{Z}\arctan(\omega\tau_z)+\sum_{s=1}^{S}\arctan\left(\frac{2\zeta_s\omega_{n_s}\omega}{\omega_{n_s}^2-\omega^2}\right)-M\times(90°)\\&\quad-\sum_{p=1}^{P}\arctan(\omega\tau_p)-\sum_{r=1}^{R}\arctan\left(\frac{2\zeta_r\omega_{n_r}\omega}{\omega_{n_r}^2-\omega^2}\right)\end{aligned} \tag{12.76}$$

其中，常数的相位为 0°，时延的幅值为 1。由这些公式可知，对应于零点和极点的每一项可以分开，先单独分析，然后再合起来构成完整的博德图（Bode diagram），即幅相频率特性曲线图。各种极点和零点的对数增益的渐近线如下：

（1）位于原点的极点

增益：$-20\lg|(j\omega)|=-20\lg\omega$。当 $\omega=1$ 时，对数增益为 0，即增益曲线在 $\omega=1$rad/s 处穿过 0 dB 点。

相位：$\Phi=-90°$。如果这类极点有多个，那么，M 重极点增益曲线变化的斜率就是 $M\times(-20)$，相位为 $M\times(-90°)$。

（2）位于实轴上的极点

增益：$-20\lg|(1+j\omega\tau_p)|=\begin{cases}0 & \omega<\frac{1}{\tau_p}\text{时}\\-20\lg(\omega\tau_p) & \omega\geq\frac{1}{\tau_p}\text{时}\end{cases}$

相位：作 $-\arctan(\omega\tau_p)$ 的渐近线，使其经过低于 $\omega=\frac{1}{\tau_p}$ 十倍频程处的 0°点以及高于 $\omega=\frac{1}{\tau_p}$ 十倍频程处的 −90°点。

极点位于 $-\frac{1}{\tau_p}$ 处。

（3）位于实轴上的零点

增益：$20\lg|1+j\omega\tau_z|=\begin{cases}0 & \omega<\frac{1}{\tau_z}\text{时}\\20\lg(\omega\tau_z) & \omega\geq\frac{1}{\tau_z}\text{时}\end{cases}$

相位：作 arctan($\omega\tau_z$) 的渐近线，使其经过低于 $\omega=\frac{1}{\tau_z}$ 十倍频程处的 0°点以及高于 $\omega=\frac{1}{\tau_z}$ 十倍频程处的 90°点。

零点位于 $-\frac{1}{\tau_z}$ 处。

（4）复数极点

增益：$-20\lg\left|1+\left(\frac{2\zeta_r}{\omega_{n_r}}\right)\mathrm{j}\omega+\left(\frac{\mathrm{j}\omega}{\omega_{n_r}}\right)^2\right|=\begin{cases}0 & \omega<\omega_{n_r}\text{时}\\ -40\lg\left(\frac{\omega}{\omega_{n_r}}\right) & \omega\geqslant\omega_{n_r}\text{时}\end{cases}$

图 12.36 的上图显示了一组实际幅频曲线，其中，$\omega_n=1.0$，ζ 取值范围为

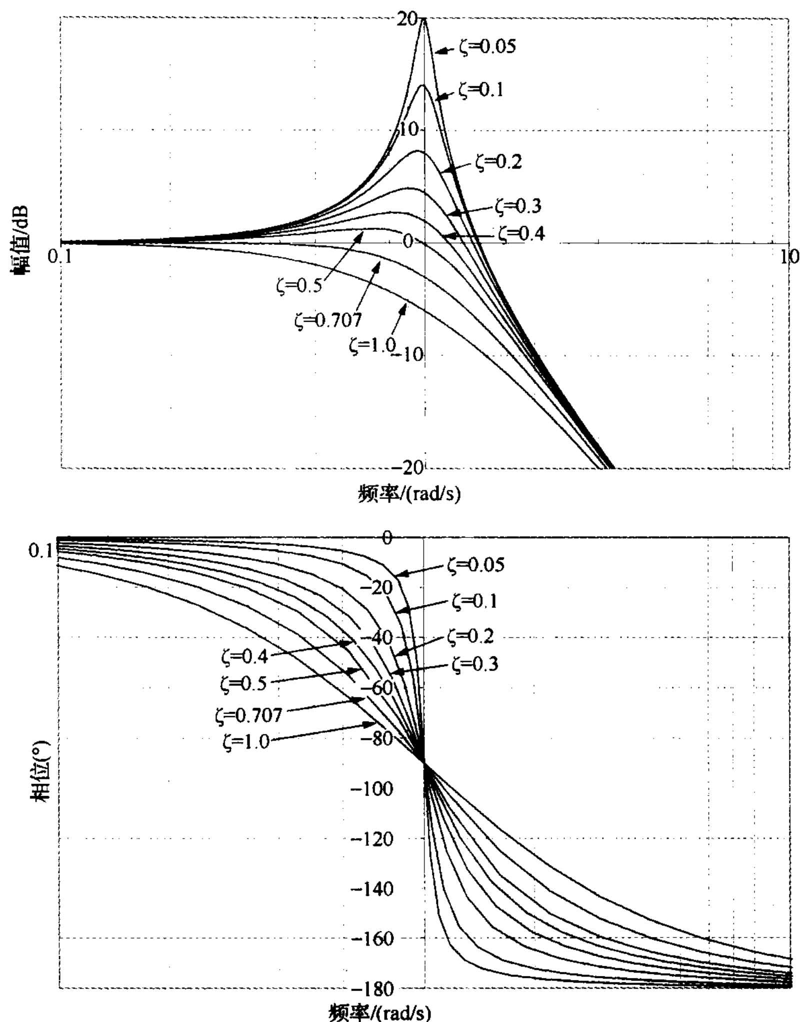


图 12.36 $\omega_n=1.0$rad/s 时的各种复数极点的博德图

0.05～1.0。注意，当ζ从1.0逐渐减小时，幅频曲线上的谐振峰逐渐增大。当ζ趋向于零时，$\omega=\omega_{n_r}$处的谐振峰的幅值趋于无穷大。$\zeta>0.707$时，则不存在谐振峰。

相位：对于不同的ζ_r值，相频曲线的变化很大，不过，一般在低于$\omega=\omega_{n_r}$十倍频程处相位为0°，在高于$\omega=\omega_{n_r}$十倍频程处相位为－180°。

图12.36的下图显示了一组实际相频曲线，其中$\omega_n=1.0$，ζ取值范围为0.05～1.0。注意，当ζ从1.0逐渐减小时，相位从0°～180°的变化速度逐渐加快，其变化过渡带所占的频率区间越来越窄。

极点位于$-\zeta_r\omega_{n_r}\pm j\omega_{n_r}\sqrt{1-\zeta_r^2}$。

(5) 复数零点

增益：$-20\lg\left|1+\left(\frac{2\zeta_s}{\omega_{n_s}}\right)j\omega+\left(\frac{j\omega}{\omega_{n_s}}\right)^2\right|=\begin{cases}0 & \omega<\omega_{n_s}\text{时}\\ 40\lg\left(\frac{\omega}{\omega_{n_s}}\right) & \omega\geqslant\omega_{n_s}\text{时}\end{cases}$

相位：对于不同的ζ_s值，相频曲线的变化很大，不过，一般在低于$\omega=\omega_{n_s}$十倍频程处相位为0°，在高于$\omega=\omega_{n_s}$十倍频程处相位为＋180°。

此时幅频和相频曲线都与图12.36所示复数极点的特性相似，只是斜率不再是－40 dB/十倍频程，而是＋40 dB/十倍频程。

零点位于$-\zeta_s\omega_{n_s}\pm j\omega_{n_s}\sqrt{1-\zeta_s^2}$。

(6) 时延

增益：对于所有ω值，增益均为1；

相位：$-\omega T_d$。

(7) 常数K

增益：$20\lg K$；

相位：0。

在博德图幅频曲线上，低频渐近线与高频渐近线相交处的频率，也就是幅频曲线斜率变化的拐点，称为转折频率。估计模型参数的第一步就是要从博德图上确定幅频曲线和相频曲线上的转折频率。在博德图上，从$-\infty$到$+\infty$，原点处的极点（或零点）具有常数斜率－20dB/十倍频程（或＋20dB/十倍频程）。实数极点（或零点）在－20dB/十倍频程（或＋20dB/十倍频程）的转折频率处存在斜率变化。复数极点（或零点）的估计要难得多，首先需要确定转折频率ω_n的位置，也就是斜率变化40dB/十倍频程的点。然后，根据实际幅频曲线上谐振峰的大小和相频曲线上180°相位变化的快慢，来估计ζ的值。

对于实轴上的极点，在转折频率处，实际的对数增益值与直线渐近线之间的误差为3 dB，也就是，该点处精确的幅频曲线值等于渐近线上的值减去3dB。距离转折频率上、下各十倍频程处，两者之间的误差减小到0.3 dB。对于实数零点，对数增益曲线与渐近线之间误差的大小也是如此，只是其精确曲线值等于渐近线加上3dB。对于复数极点或零点，在转折频率处，实际对数增益与渐近线之间的误差与ζ的值相

关，由图 12.36 可见，两者之间的误差可以很大。

例 12.17 假设在各个不同频率正弦波的输入驱动下，某开环系统的输出测量数据如下所示。请画出其博德图，并估计其传递函数。

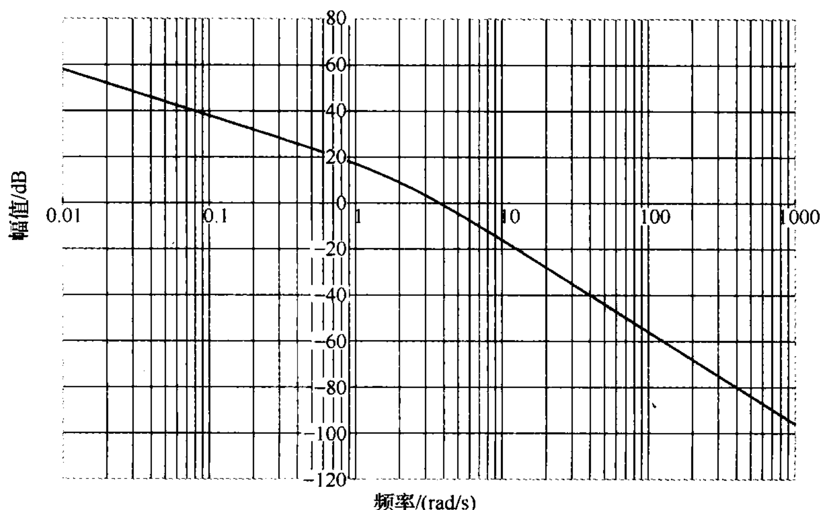
频率/(rad/s)	0.01	0.02	0.05	0.11	0.24	0.53	1.17	2.6	5.7	12.7	28.1	62	137	304	453	672	1000
20lg|G|/dB	58	51	44	37	30	23	15	6	−7	−20	−34	−48	−61	−75	−82	−89	−96
相位(°)	−90	−91	−91	−93	−97	−105	−120	−142	−161	−171	−176	−178	−179	−180	−180	−180	−180

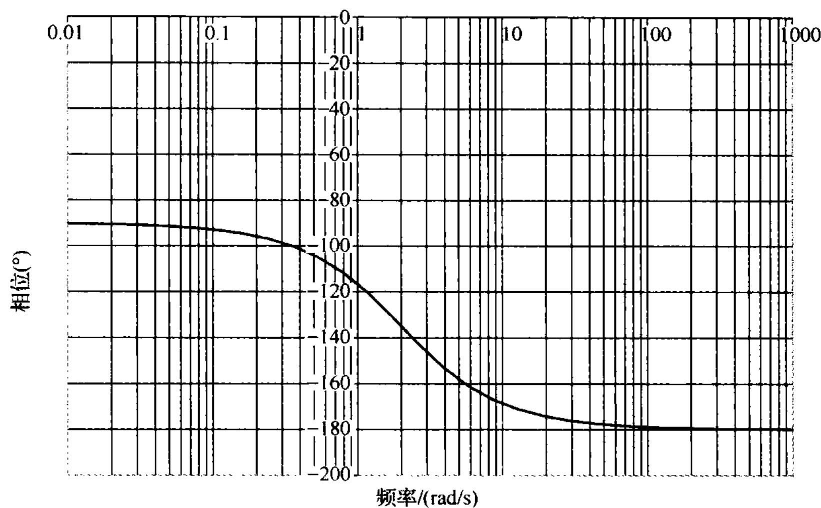
解：根据给定数据所作的幅频曲线和相频曲线博德图如下图所示。由相频曲线可见，当 $\omega\to\infty$时，相位趋向于 180°，因此，该系统的极点应该比零点多 2 个。另外，注意，幅频曲线上没有谐振峰，相频曲线的变化也较缓慢，因此，系统中没有小阻尼（即 $\zeta<0.5$）复数极点。不过，现在还不能确定是否存在大阻尼复数极点。

在 0.01～1rad/s 频率区间，幅频曲线的斜率为－20 dB/十倍频程。不妨假设小于 0.01rad/s 的频率范围也具有同样的斜率，那么，就可以确定系统传递函数在原点处有一个极点，这样，才会产生如此的响应曲线。如果可能，在记录仪器允许范围内，必须验证小于 0.01rad/s 的频率下幅频响应是否确实保持－20 dB/十倍频程的斜率。

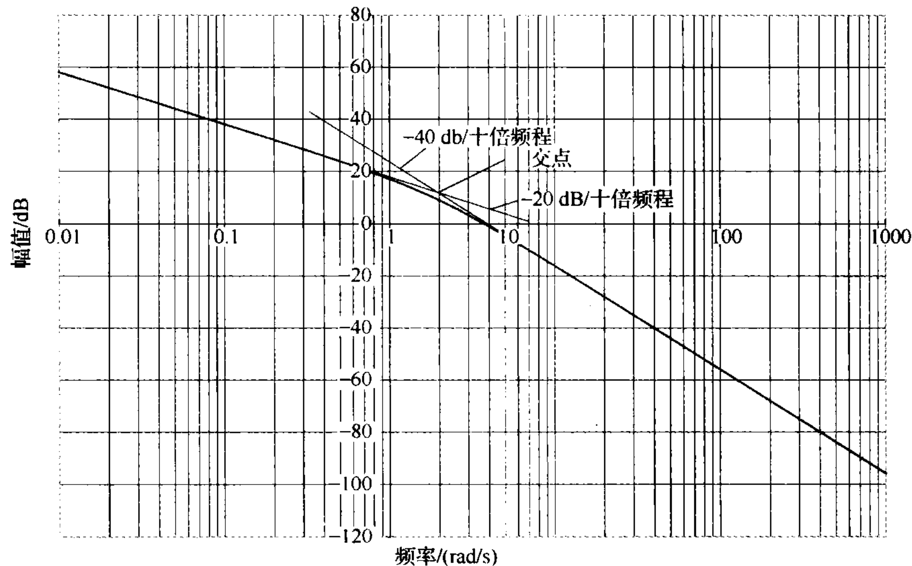
在 2～1000rad/s 频率区间，幅频曲线的斜率为－40 dB/十倍频程。由于在其之前那段频率区间已经确认了一个极点，因此，可以认为在 2～1000rad/s 内存在另一个极点。当然，也有其他可能性，如存在相距很近的一个极点和一个零点，或者存在一对复数极点和一个零点。但是，这些可能性的证据不足，为了简便，这里用一个极点来描述系统就足够了。

至此，已经可以认定系统中含有一个原点处的极点和一个位于 1rad/s 之上的极点。而且，整个幅频曲线的斜率已经完成匹配，因此，现在可以确定系统中不存在大阻尼复数极点。





为了估计极点的位置，如下图所示，作低频和高频两条幅频曲线的切线。两条切线的交点约在 2rad/s 处，这就是转折频率，也就是极点的位置。注意，如前所述，从图中可见，转折频率处实际曲线约比渐近线低 3 dB。



至此，可以写出模型的传递函数为

$$G(\mathrm{j}\omega)=\frac{K}{\mathrm{j}\omega\left(\frac{\mathrm{j}\omega}{2}+1\right)}$$

其中，常数 K 可以通过 $\omega=1\mathrm{rad/s}$ 处的幅值来计算。此处，原点处的极点对于对数增益的贡献为 0，而 2rad/s 处的极点对于对数增益的贡献为－1 dB，即$\left(-20\lg\left|\frac{j\omega}{2}+1\right|=\right.$

$-20\lg\left(\sqrt{\left(\frac{1}{2}\right)^2+1^2}\right)=-1\text{dB}\right)$。由于幅频曲线上的 1rad/s 点处于转折频率上、下十倍频程的范围之内，因此，需要计算位于 2rad/s 处的极点的贡献。于是，$\omega=1\text{rad/s}$ 处的非零增益项就是

$$20\lg|G(\omega)|=20\lg K-20\lg\left|\frac{j\omega}{2}+1\right|$$

由以上幅频曲线可知：$20\lg|G(\omega)|=17$ dB。因此

$$17=20\lg K-1$$

即 $K=8$。这样，模型方程就成为

$$G(j\omega)=\frac{8}{j\omega\left(\frac{j\omega}{2}+1\right)}$$

最后还要检查系统是否存在时延。在 $\omega=2.0\text{rad/s}$ 转折频率处，由以上模型方程求得的相位为

$$\phi(\omega)=-90-\arctan\left(\frac{\omega}{2}\right)=-90-45=-135°$$

该数值与相频曲线上的数据很接近，因此，可以认为系统中并没有时延。■

以上例 12.17 说明了确定模型结构和参数的方法。但是，用这种方法分析复杂系统，即便不是完全不可能，也是非常难的。不过，很多软件包可以根据所提供的数据自动估计系统模型的极点、零点、时延以及传递函数的增益等参数，如 MATLAB 中的系统辨识工具箱等。还有其他比 MATLAB 更灵活的程序，如 Seidel 在 1975 完成的 FORTRAN 程序等，也可以用于分析复杂系统。

除了正弦波以外，还可以用其他输入信号来激励系统产生输出响应，例如：脉冲、阶跃等瞬变信号和斜坡信号，以及白噪声和伪随机二进制序列等噪声信号。之所以要用其他信号，是因为并不是所有系统都可以用正弦信号来激励，本章所述的快速眼动系统就是其中一例，对于这类系统，我们一般使用阶跃输入信号。

利用阶跃输入的输出响应数据分析系统传递函数时，通常使用傅里叶变换和快速傅里叶变换（FFT）将时域响应转化为频域响应，用频域法来分析。输入信号的频域响应是已知的（如 s^{-1}），输出信号的频域响应则可以用数值方法（即 FFT）计算得到。不过，必须先用 A/D 转换器将系统输出信号进行数字化采样，存入计算机内存之后，才能进行 FFT 运算。注意，A/D 采样之前，必须先用抗混叠滤波器（即低通滤波器）滤除信号最高频率以上的成分，并且所用的采样率至少是信号最高频率的 2.5 倍。输入和输出信号的频域响应得到之后，就可以根据式（12.67）计算系统的传递函数。

12.10.2　一阶线性系统的辨识

下面介绍一种专用于一阶系统的时域辨识方法。设某系统为一阶系统，其输入激励为阶跃信号，那么，如图 21.37 所示，系统的输出响应就是

$$v_0(t)=v_{ss}+Ke^{-\frac{t}{\tau}}u(t) \tag{12.77}$$

其中，$K=-(v_{ss}-v(0))$，τ 为时间常数，$v(0)$ 为初始值，v_{ss}为稳态值。注意，当 $t=\tau$ 时，输出响应已达到初始值与稳态值之差的 63%；同理，当 $t=4\tau$ 时，输出响应达到初始值与稳态值之差的 98%。

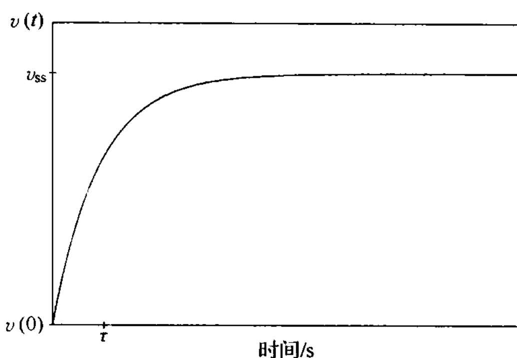


图 12.37 一阶系统阶跃输入的输出响应

为了确定该一阶系统模型，需要估计式（12.77）中的几个参数。如图 12.38 所示，系统时间常数 τ 的一种估计方法是求输出响应曲线的初始斜率，并用平均法求稳态值。即

$$\dot{v}(t)=\frac{1}{\tau}(v_{ss}-v(0))\mathrm{e}^{-\frac{t}{\tau}}\rightarrow\tau=\frac{v_{ss}-v(0)}{\dot{v}(t)}\mathrm{e}^{-\frac{t}{\tau}}\qquad(12.78)$$

$t=0$ 时，$\tau=\frac{v_{ss}-v(0)}{\dot{v}(0)}$，其中 $\dot{v}(0)$ 是输出响应的初始斜率。以上求 τ 的方程其实是一个直线方程。

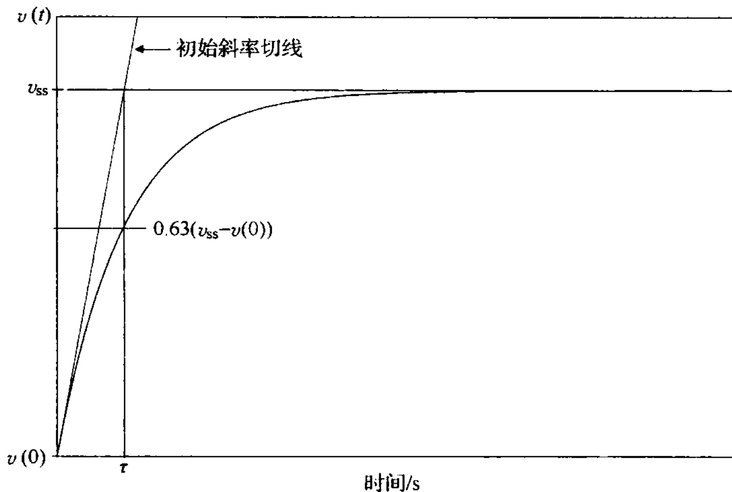


图 12.38 由系统输出响应的初始斜率计算时间常数

例 12.18 某一阶系统的阶跃响应如下，请计算系统模型的各个参数。

t	0	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5	0.7	1.0	1.5	2.0
$v(t)$	0.0	0.56	0.98	1.30	1.54	1.73	1.86	2.04	2.15	2.24	2.27	2.28	2.28

解：该系统模型可以用式（12.77）来描述，即

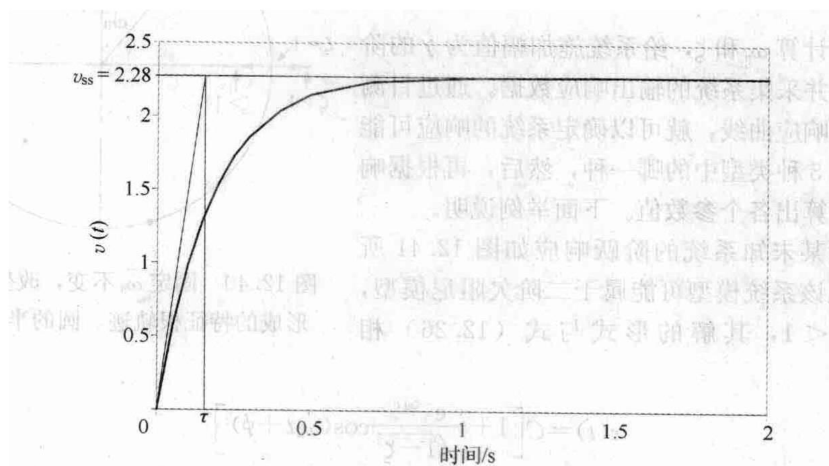
$$v(t)=v_{ss}+Ke^{-\frac{t}{\tau}}u(t)$$

其中，v_{ss}、K 和 τ 为未知参数。显然，$v_{ss}=2.28$，$K=-2.28$。下图所示是阶跃响应曲线，以及斜率为 $\dot{v}(0)$ 的切线。由图中可见，$\tau=1.7\text{s}$。因此，模型方程可以写为

$$v(t)=2.28(1-e^{-\frac{t}{1.7}})u(t)$$

其传递函数即为

$$V(s)=\frac{v_{ss}}{s\tau+1}=\frac{2.28}{0.17s+1}=\frac{13.4}{s+5.9}$$



12.10.3　二阶线性系统的辨识

下面用时域法估计二阶线性系统的参数。

二阶线性系统的微分方程为 $f(t)=M\ddot{x}+B\dot{x}+Kx$，为了便于分析，将此方程写成如下标准形式

$$\frac{f(t)}{M}=\ddot{x}+2\zeta\omega_n\dot{x}+\omega_n^2x \tag{12.79}$$

式中　ω_n——无阻尼自然频率；

ζ——阻尼比。

对于图 12.39 所示的系统，这两个参数分别为

$$\zeta=\frac{B}{2\sqrt{KM}},\quad \omega_n=\sqrt{\frac{K}{M}}$$

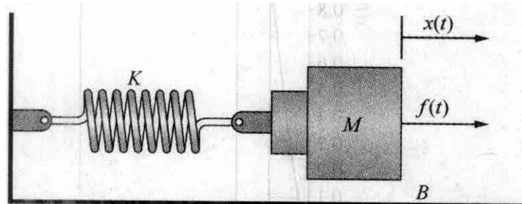


图 12.39　简单机械系统

特征方程的根为

$$s_{1,2}=-\zeta\omega_n\pm\omega_n\sqrt{\zeta^2-1}=-\zeta\omega_n\pm j\omega_n\sqrt{1-\zeta^2}=-\zeta\omega_n\pm j\omega_d \tag{12.80}$$

固定 ω_n 的值不变，改变 ζ 的值，如图 12.40 所示，这两个特征根会在复平面上移动。当 $0<\zeta<1$ 时，系统为欠阻尼系统；$\zeta=1$ 时，系统为临界阻尼系统；而 $\zeta>1$ 时，系统为过阻尼系统。应用经典方法，即通过求解特征方程的根，可以分别得到这 3 种系统的自由解（即齐次解）如下：

系统的阻尼　　　　自由解

过阻尼　$x(t)=x_{ss}+A_1e^{s_1t}+A_2e^{s_2t}$

欠阻尼 $x(t)=x_{ss}+B_1 e^{-\zeta\omega_n t}\cos(\omega_d t+\phi)$

临界阻尼 $x(t)=x_{ss}+(C_1+C_2 t)e^{-\zeta\omega_n t}$

其中 x_{ss} 是 $x(t)$ 的稳态值，A_1、A_2、B_1、ϕ、C_1 和 C_2 为系统常数，它们由系统的初始条件决定。

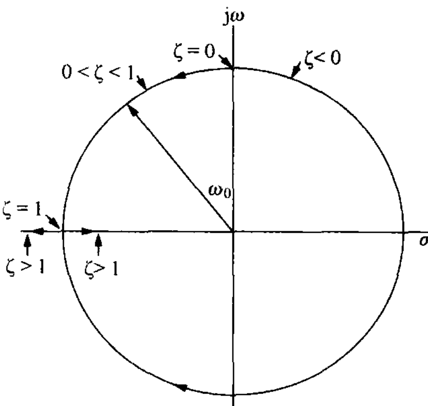


图 12.40 固定 ω_n 不变，改变 ζ 时所形成的特征根轨迹。圆的半径为 ω_n

为了计算 ω_n 和 ζ，给系统施加幅值为 γ 的阶跃输入，并采集系统的输出响应数据。通过目测观察输出响应曲线，就可以确定系统的响应可能属于以上 3 种类型中的哪一种，然后，再根据响应曲线计算出各个参数值。下面举例说明。

假设某未知系统的阶跃响应如图 12.41 所示。可见该系统模型可能属于二阶欠阻尼模型，即 $0<\zeta<1$，其解的形式与式（12.26）相似，为

$$x(t)=C\left[1+\frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}}\cos(\omega_d t+\phi)\right] \tag{12.81}$$

式中 C——稳态响应幅值 x_{ss}。

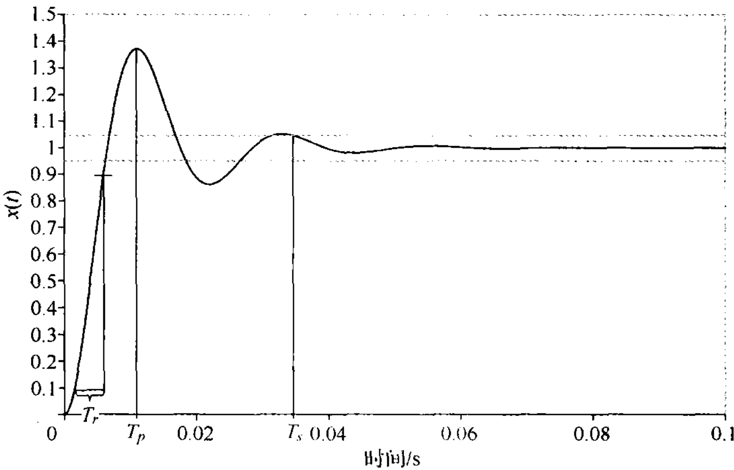


图 12.41 某二阶系统的阶跃响应

注：图中标出了上升时间 T_r、峰值时间 T_p 和调节时间 T_s。

一般用以下几个时间参数可以定量描述阶跃输入的输出响应（见图 12.41）：

上升时间 T_r 为响应曲线从稳态值的 10% 上升到 90% 所需要的时间；调节时间 T_s 为响应曲线在此 T_s 时间之后就一直保持在稳态值上下 5% 的范围之内；峰值时间 T_p 为响应曲线达到第一个过冲峰值的时间。

如图 12.41 所示，利用响应曲线上可以确定的这几个参数值就可以建立系统模

型。由例 12.10 可知，二阶欠阻尼系统的峰值时间计算公式为

$$T_p=\frac{\pi}{\omega_n\sqrt{1-\zeta^2}} \tag{12.82}$$

T_p时刻的输出响应为

$$x(T_p)=C\left(1+e^{\zeta\frac{\pi}{\sqrt{1-\zeta^2}}}\right) \tag{12.83}$$

根据响应曲线上测得的 T_p和 x（T_p）的值，就可以利用以上两个公式计算系统的 ζ 和 ω_n值。首先用式（12.83）计算 ζ 值；然后，再将算得的 ζ 值代入式（12.82），计算 ω_n值；最后，用 ζ 值和式（12.26），计算式（12.81）中的相位 ϕ，即

$$\phi=\pi+\arctan\frac{-\zeta}{\sqrt{1-\zeta^2}}$$

例 12.19　请根据图 12.41 所示数据计算系统模型的 ζ 和 ω_n。

解： 由图 12.41 所示数据可知，$C=1.0$，$T_p=0.011$，x（T_p）$=1.37$。因此

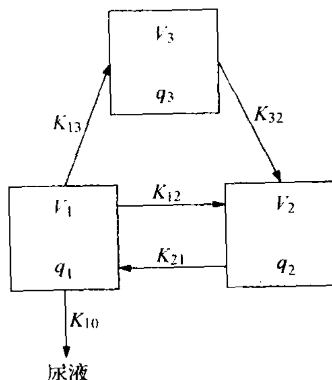
$$x(T_p)=C\left(1+e^{\zeta\frac{\pi}{\sqrt{1-\zeta^2}}}\right)\rightarrow\zeta=\sqrt{\frac{\frac{(\ln(x(T_p)-1))^2}{\pi^2}}{1+\frac{(\ln(x(T_p)-1))^2}{\pi^2}}}=0.3$$

$$T_p=\frac{\pi}{\omega_n\sqrt{1-\zeta^2}}\rightarrow\omega_n=\frac{\pi}{T_p\sqrt{1-\zeta^2}}=300\text{rad/s}$$ ■

习　　题

1. 将 500mg 染料注入血浆房室，达到稳态之后，血液中的染料浓度为 0.0893mg/cm^3。请计算血浆房室的容积。

2. 某房室系统如下图所示，将一种放射性标记过的溶质快速注入房室 II。请分别写出只包含变量（a）q_1；（b）q_2；（c）q_3的单一微分方程。



3. 有两个体积均为 0.0572cm^3 的房室，房室之间由一层薄膜隔开。如果开始时

全部溶质都注入其中一个房室，请计算溶质的传输速率。设时间常数为 $27\times10^3 s^{-1}$。

4. 假设房室 I 的容积为 $0.0572cm^3$，比房室 II 大 1 倍。如果将 100mol 某种溶质注入房室 II 中，求两个房室中该溶质的浓度。

5. 将某种放射性标记过的物质通过静脉注射到血浆中。血浆中该物质随时间变化的浓度遵循如下关系式

$$c_1 = A_1 e^{\lambda_1 t} + A_2 e^{\lambda_2 t}$$

式中 c_1——血浆中的物质浓度，单位为 mg/100ml；

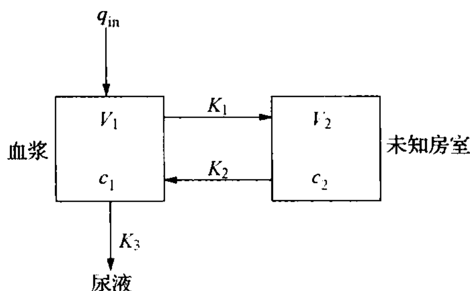
A_1——143mg/100ml；

A_2——57mg/100ml；

λ_1——-1.6 天$^{-1}$；

λ_2——-2.8 天$^{-1}$。

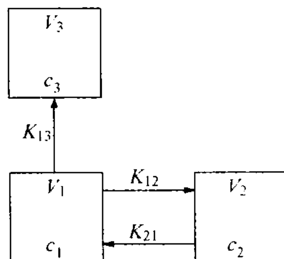
物质的注射量为 10mg。系统的房室模型如下



(a) 请计算血浆房室的容积；(b) 请计算传输速率 K_1、K_2和 K_3。

6. 假设 15mg/s 大小的阶跃输入作用下，染料稀释试验的阶跃响应稳态值为 $15/70mg/cm^3$，时间常数为 10/7s。(a) 请计算心输出量和心肺房室的容积；(b) 假设输入是持续时间为 1s 的 15mg/s 脉冲信号，请计算出心肺房室的染料浓度。

7. 设三房室模型如下



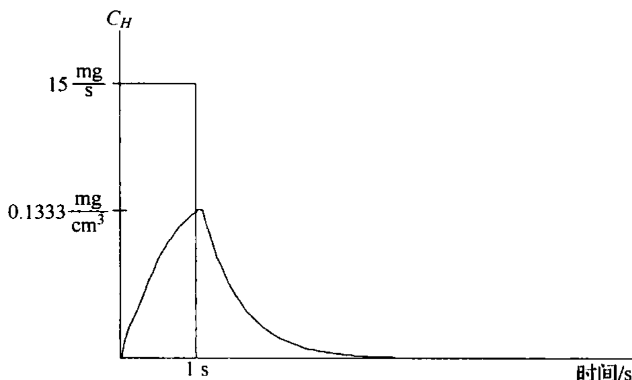
将 10mg 放射性标记溶质快速团注输入房室 I 之后，房室 I 中溶质浓度随时间变化的关系式为

$$c_1 = 57e^{-2.8t} + 143e^{-1.6t}\frac{mg}{100ml}$$

其中 t 的单位为天。(a) 请问房室 I 的容积是多大？(b) 请计算传输速率 K_{12}、K_{13}

和 K_{21}。

8. 下图所示是染料稀释试验的脉冲响应数据，为了便于观察，输入信号和浓度曲线重叠显示在一起。已知脉冲宽度小于 4 倍时间常数，并且 $\int_0^\infty c_H \mathrm{d}t = 0.2\ \frac{\mathrm{mg \cdot s}}{\mathrm{cm}^3}$。请计算心输出量和心肺房室的容积。



9. 放射性物质 I^{131} 被团注输入血浆房室。血浆中 I^{131} 浓度随时间变化的关系式为 $c_1 = 143e^{-1.6t}\ \frac{\mathrm{mg}}{100\mathrm{ml}}$。$I^{131}$ 的注射量为 10mg。假设血浆的房室模型与习题 12.1 相同，请计算（a）血浆房室的容积；（b）$K = K_1 + K_2$。

10. 将质量未知的放射性碘同位素 I^{131} 瞬时注入血浆中。血浆中 I^{131} 的量随时间以指数函数的形式逐渐从 100mg 开始下降，其时间常数为 1 天；而尿液中 I^{131} 的量也以指数函数的形式逐渐从 0 上升到 75mg，时间常数也是 1 天。假设房室模型与习题 12.1 相同，请计算 K_1 和 K_2 的值。

11. 什么是眼动系统的主特性曲线？请将本章 12.4 节讲述的 Westheimer 模型所得的仿真结果与各个主特性曲线进行比较。

12. 请设计 SIMULINK 仿真程序，用本章 12.4 节所述的 Westheimer 模型仿真 20°扫视眼动过程，其中 $\zeta = 0.7$，$\omega_n = 120\mathrm{rad/s}$，且设 $K = 1\mathrm{Nm}$。并对 5°、10°、15°扫视进行同样的仿真，将这些仿真结果与图 12.11 所示的主特性曲线进行比较。

13. 假设 12.4 节中 Westheimer 模型的输入是 12.5 节所述的脉冲一阶跃信号，并且 $\zeta = 0.7$，$\omega_n = 120\mathrm{rad/s}$，$K = 1\mathrm{Nm}$。（a）请计算能够使眼球保持在 20°所需的阶跃输入信号的大小；（b）设计 SIMULINK 仿真程序，计算能够驱动眼球完成 20°扫视的脉冲幅值，并要求仿真结果与图 12.11 所示的主特性曲线相匹配；（c）用（b）中设计的程序仿真 5°、10°和 15°扫视；（d）将这些仿真结果与 Westheimer 模型的仿真结果进行比较。

14. 某个扫视眼动系统的模型可以用如下方程表示

$$\tau = 1.74 \times 10^{-3}\ \ddot{\theta} + 0.295\ \dot{\theta} + 25\theta$$

其中 τ 是所施加的力矩。假设 $\tau = 250u(t)$，且初始条件为 0。请应用拉普拉斯变换求

解 $\theta(t)$，并作出 $\theta(t)$ 曲线。

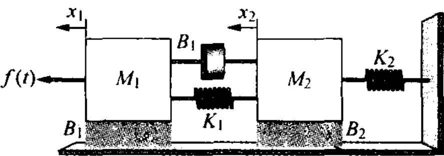
15. 假设某病人口服了少量放射性碘同位素 I^{131}，并且，I^{131} 从血流排入尿液和甲状腺的过程可以用简单的一阶微分方程来描述。设 I^{131} 从血流进入甲状腺和尿液的传输速率分别为 K_1 和 K_2，系统的时间常数为 $\tau=\frac{1}{K_1+K_2}$。(a) 请作出系统的响应曲线；(b) 假设甲状腺不工作，不吸收 I^{131}，请作出此时的系统响应曲线，并将其与 (a) 中的结果进行比较。

16. 设某地区人口中有易感染者 50000 名，已感染者 10 名，获免疫者为 0，且 $\beta=0.001$，$\gamma=0.8$，请应用 Kermack-McKendrick 模型仿真此传染病的流行曲线。

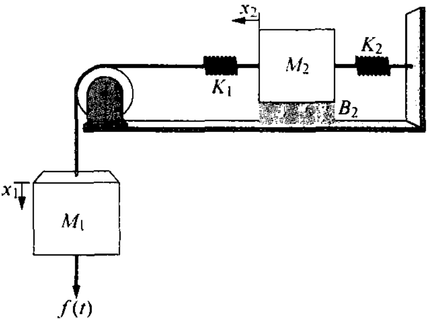
17. 设某地区人口中有易感染者 50000 名，已感染者 10 名，获免疫者为 0，且 $\beta=0.001$，$\gamma=0.8$，$\delta=0.01$。在考虑出生率和死亡率的情况下，请设计 SIMULINK 仿真程序，用 Kermack-McKendrick 模型仿真此传染病的流行曲线。

18. 设某地区人口中有易感染者 50000 名，已感染者 20 名，获免疫者为 0，且 $q=0.98896$。请设计 SIMULINK 仿真程序，用 Reed-Frost 模型求此传染病的流行曲线。

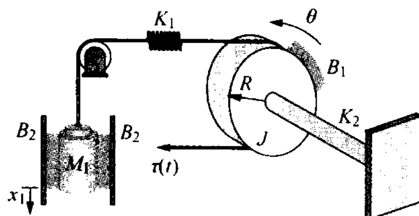
19. 下图所示系统中：$M_1=2\text{kg}$，$M_2=1\text{kg}$，$B_1=1\ (\text{N}\cdot\text{s})/\text{m}$，$B_2=2(\text{N}\cdot\text{s})/\text{m}$，$K_1=1\text{N/m}$，$K_2=1\text{N/m}$。且 $f(t)$ 为外加作用力，x_1 和 x_2 为偏离静止位置的位移。(a) 求系统的传递函数；(b) 请编写 MATLAB 程序，作出系统的博德图。



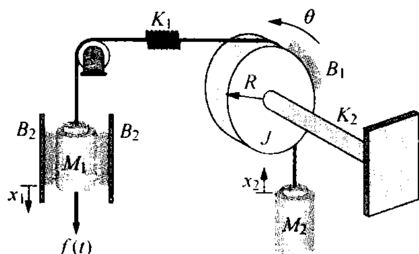
20. 下图所示系统中：$M_1=1\text{kg}$，$M_2=2\text{kg}$，$B_2=2\ (\text{N}\cdot\text{s})/\text{m}$，$K_1=1\text{N/m}$，$K_2=1\text{N/m}$。且 $f(t)$ 为外加作用力，x_1 和 x_2 为偏离静止位置的位移。假设滑轮没有惯性和摩擦。(a) 请写出系统模型的微分方程组；(b) 如果输入信号是幅值为 10N 的阶跃函数，请设计 SIMULINK 程序，仿真位移 x_1 和 x_2 的响应曲线。



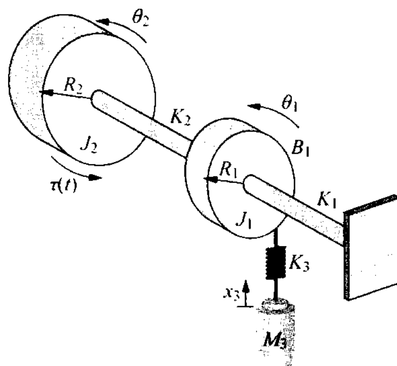
21. 下图所示系统包括粘性元件 B_2、平移元件 K_1 和转动元件 K_2。设 $\tau(t)$ 为外加力矩，x_1 为质量 M 偏离静止位置的位移，θ 为元件 J 偏离静止位置的角位移。元件 J 处于静止位置时，弹簧处于没有任何伸缩的自然状态。(a) 写出描述该系统的微分方程组；(b) 写出描述该系统的状态变量方程组。



22. 下图所示系统包括两个粘性元件 B_1 和 B_2、平移元件 K_1 和转动元件 K_2。设 $f(t)$ 为外加作用力，x_1 和 x_2 分别为质量 M_1 和 M_2 偏离静止位置的位移，θ 为元件 J 偏离静止位置的角位移。元件 J 处于静止位置时，弹簧处于没有任何伸缩的自然状态。假设滑轮没有惯性和摩擦，缆绳不会被拉长。(a) 写出描述该系统的微分方程组；(b) 写出描述该系统的状态变量方程组。



23. 下图所示系统中：$M_3=0.5\text{kg}$，$K_3=4\text{N/s}$（平移运动），$J_1=0.5\text{kg}\cdot\text{m}^2$，$B_1=1\text{N}\cdot\text{s}\cdot\text{m}$，$K_1=2\text{N}\cdot\text{m}$，$J_2=2\text{kg}\cdot\text{m}^2$，$K_2=1\text{N}\cdot\text{m}$，$R_1=0.2\text{m}$，$R_2=1.0\text{m}$。设 $\tau(t)$ 为外加力矩，x_3 为质量 M_3 偏离静止位置的位移，θ_1 和 θ_2 为元件 J_1 和 J_2 偏离静止位置的角位移。元件 J_1 和 J_2 处于静止位置时，弹簧没有任何拉伸。(a) 写出描述该系统的微分方程组；(b) 如果输入信号是幅值为 10N 的阶跃函数，请设计 SIMULINK 程序，仿真角位移 θ_1、θ_2 和线性位移 x_3 的响应曲线。



24. 设本章 12.4 节所述 Westheimer 模型的参数为 $\zeta=\frac{1}{\sqrt{2}}$，$\omega_n=100\text{rad/s}$，$K=1$，求出如图 12.11 所示那样的脉冲一阶跃输入信号的输出响应通式。并考察脉冲幅值增加、持续时间减少、但眼动扫视的稳态幅度保持不变的情况下，该模型输出响应的变化。

25. 请利用图 12.11 所示的主特性曲线数据，分别估计 5°、10°、15°和 20°扫视下 Westheimer 模型（见 12.4 节）的参数 ζ 和 ω_n 的值。设扫视的最大过冲幅值 $x(T_p)$ 比扫视稳态幅度大 1°。请仿真这 4 个角度的扫视过程；并建立随扫视幅度变化的参数 ζ 与 ω_n 之间的关系式，使其与图 12.11 的主特性曲线相匹配。然后，根据这些关系，分别作出 T_p 和扫视峰速随扫视幅度变化的函数曲线，并将这些曲线与原 Westheimer 模型的主特性曲线以及图 12.11 所示曲线进行比较。

26. 设处于未激活状态的眼肌的模型如图 12.28 所示，其中 $K_{lt}=32\text{N/m}$，$K_{se}=125\text{N/m}$，$B=3.4\text{Ns/m}$。且对于未激活肌肉，$F=0$。（a）求模型的传递函数 $H(j\omega)=\frac{X_1}{T}$；（b）编写 MATLAB 程序，作出模型的博德图。

27. 设处于未激活状态的眼肌的模型如图 12.29 所示，其中 $K_{lt}=60.7\text{N/m}$，$K_{se}=125\text{N/m}$，$B_1=2\text{Ns/m}$，$B_2=0.5\text{Ns/m}$。且对于未激活肌肉，$F=0$。（a）求模型的传递函数 $H(j\omega)=\frac{X_1}{T}$；（b）编写 MATLAB 程序，作出模型的博德图。

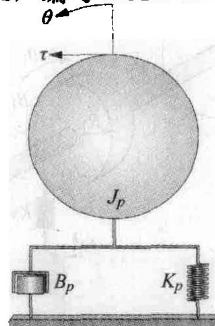
28. 请设计 SIMULINK 程序，用 12.8 节的线性同胚扫视眼动模型，仿真 20°扫视过程。

29. 请设计 SIMULINK 程序，用 12.9 节的线性同胚扫视眼动模型，仿真 20°扫视过程。

30. 给定式（12.41）表示的线性同胚扫视眼动模型。（a）写出其传递函数；（b）编写 MATLAB 程序，作出模型的博德图。

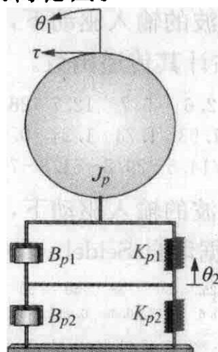
31. 请验证图 12.30 所示的各条长度—张力函数曲线。

32. 下图所示是眼眶组织的被动特性模型，设其受到的驱动力矩为 τ，且 $K_p=0.5\text{g/(°)}$，$B_p=0.06\text{gs/(°)}$，$J_p=4.3\times10^{-5}\text{gs}^2\text{/(°)}$。所有元件均为转动元件。（a）求模型的传递函数 $\theta(s)/\tau(s)$；（b）编写 MATLAB 程序，作出模型的博德图。

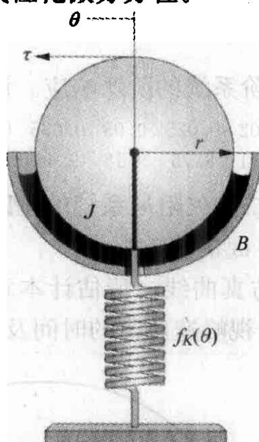


33. 请验证图 12.31 所示肌肉模型的力一速度函数曲线。提示：设计 SIMULINK 程序，计算各个不同 M 值所对应的峰速。

34. 下图所示是眼眶组织的被动模型，设其受到的驱动力矩为 τ，且 $J_p=4.308\times10^{-5}\ \text{gs}^2/(°)$，$K_{p1}=0.5267\text{g}/(°)$，$K_{p2}=0.8133\text{g}/(°)$，$B_{p1}=0.010534\text{gs}/(°)$，$B_{p2}=0.8133\text{gs}/(°)$。所有元件均为转动元件。(a) 求模型的传递函数 $\theta_1(s)/\tau(s)$；(b) 用 MATLAB 程序作出模型的博德图。



35. 下图所示是眼动系统模型。设所有元件均为转动元件，且 $f_K(\theta)=K_1\theta^2$（非线性转动弹簧的方程）。(a) 写出该系统的非线性微分方程；(b) 利用泰勒级数一阶近似法，写出某工作点附近的线性化微分方程。



36. 设未激活肌肉的被动弹性模型可以用如下非线性力一位移的平移方程表示

$$f(x)=|x|x$$

其中 x 是偏离平衡位置的位移。请计算该非线性元件在平衡位置附近的线性逼近。

37. 假设在各个不同频率正弦波的输入驱动下，某开环系统的输出测量数据如下所示。请画出系统的博德图，并估计其传递函数。

频率/(rad/s)	0.6	1.6	2.6	3.6	5.5	6.1	7.3	9.8	12.7	32.9	62.1	100
输出与输入幅值比 $\|G\|$	2.01	2.03	2.09	2.17	2.37	2.43	2.49	2.16	1.39	0.18	0.05	0.018
相位(°)	−3.3	−8.9	−15.0	−21.8	−38.1	−44.8	−60.0	−93.6	−123.2	−164.2	−170.0	−175.0

38. 假设在各个不同频率正弦波的输入驱动下，某开环系统的输出测量数据如下所示。请画出系统的博德图，并估计其传递函数。

频率/(rad/s)	0.001	0.356	1.17	2.59	8.53	12.7	18.9	41.8	62.1	137	304	1000
20lg | G | /dB	6.02	6.02	5.96	5.74	3.65	1.85	−0.571	−6.64	−9.95	−16.8	−23.6	−34.0
相位(°)	−0.086	−3.06	−10.0	−22.0	−64.9	−88.1	−116.0	−196.0	−259.0	−479.0	−959.0	−2950

39. 假设在各个不同频率正弦波的输入驱动下，某开环系统的输出测量数据如下所示。请画出系统的博德图，并估计其传递函数。

频率/(rad/s)	0.11	0.24	0.53	1.17	2.6	5.7	12.7	28.1	62	137	304	453	672	1000
幅值比 | G |	2.0	2.0	2.0	2.0	1.93	1.74	1.24	0.67	0.32	0.15	0.07	0.044	0.03	0.02
相位(°)	−0.62	−1.37	−3.03	−6.7	−14.5	−29.8	−51.8	−70.4	−80.9	−85.8	−88.1	−88.7	−89.1	−89.4

40. 假设在各个不同频率正弦波的输入驱动下，某开环系统的输出测量数据如下所示。请画出系统的博德图。(数据取自 Seidel，1975。)

频率/(rad/s)	1	3	7	10	15	20	25	30	35	40	50	60	70	80	90	100	110	120	130	140
幅值比 | G |	1	0.95	0.77	0.7	0.67	0.63	0.6	0.53	0.48	0.44	0.35	0.31	0.33	0.35	0.32	0.32	0.3	0.29	0.27	0.26
相位(°)	−0.35	−.227	−.419	−.541	−.611	−.768	.995	−1.08	1.24	−1.31	−1.52	−1.92	−1.61	−1.83	−2.08	−2.23	−2.53	−2.72	−2.9	−3.0

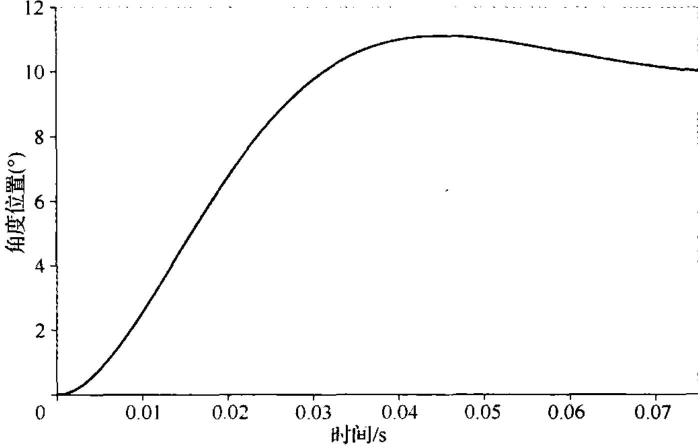
请在如下各种情况下估计系统的传递函数：(a) 2 个实数极点；(b) 1 个实数极点和 1 对复数极点；(c) 2 个实数极点，1 个零点和一对复数极点；(d) 3 个实数极点，1 个零点和一对复数极点。提示：可以利用 MATLAB 的系统辨识工具箱或者 Seidel 程序来求解此题。

41. 下列数据是某个未知一阶系统的阶跃响应。请求出该一阶模型的各个参数。

T	0.0	0.005	0.01	0.015	0.02	0.025	0.03	0.035	0.04	0.045	0.05	0.055	0.06	0.1
$v(t)$	0.00	3.41	5.65	7.13	8.11	8.75	9.18	9.46	9.64	9.76	9.84	9.90	9.93	10.

42. 假设方程 12.81 是某个二阶欠阻尼系统的阶跃响应，且 $C=10$，$T_p=0.050$，$x(T_p)=10.1$。请计算参数 ζ 和 ω_n 的值。

43. 下图所示是 10°扫视的仿真曲线。请估计本章 12.4 节所述 Westheimer 模型的 ζ 和 ω_n 参数值，并计算达到扫视峰速所需的时间及其峰速。



推荐阅读资料

Ackerman, E. and Gatewood, L.C. (1979). *Mathematical Models in the Health Sciences: A Computer-aided Approach*. Univ. Minnesota Press, Minneapolis.

Bahill, A.T., Latimer, J.R. and Troost, B.T. (1980). Linear homeomorphic model for human movement. *IEEE Trans. Biomed. Eng.* **27**, 631–639.

Bahill, A.T. (1981). *Bioengineering: Biomedical, Medical and Clinical Engineering*. Prentice-Hall, Englewood Cliffs, NJ.

Carpenter, R.H.H. (1988). *Movements of the Eyes*. Pion, London.

Carson, E. and Cobelli, C. (2001). *Modeling Methodology for Physiology and Medicine*. Academic, San Diego.

Close, R.I. and Luff, A.R. (1974). Dynamic properties of inferior rectus muscle of the rat. *J. Physiol.* **236,** 259–270.

Collins, C.C. (1975). The human oculomotor control system. In *Basic Mechanisms of Ocular Motility and Their Clinical Implications* (G. Lennerstrand and P. Bach-y-Rita, Eds.). Pergamon, Oxford, 145–180.

Collins, C.C., O'Meara, D., and Scott, A.B. (1975). Muscle tension during unrestrained human eye movements. *J. Physiol.* **245**: 351–369.

Enderle, J.D., Wolfe, J.W. and Yates, J.T. (1984). The linear homeomorphic saccadic eye movement model—A modification. *IEEE Trans. Biomed. Eng.* **31(11)** 717–720.

Enderle, J.D. and Wolfe, J.W. (1987). Time-optimal control of saccadic eye movements. *IEEE Trans. Biomed. Eng.* **BME-34(1):** 43–55.

Enderle, J.D. and Wolfe, J.W. (1988). Frequency response analysis of human saccadic eye movements: estimation of stochastic muscle forces. *Computers in Biol. & Med.* **18(3),** 195–219.

Enderle, J.D., Engelken, E.J. and Stiles, R.N. (1991). A comparison of static and dynamic characteristics between rectus eye muscle and linear muscle model predictions. *IEEE Trans. Biomed. Eng.* **BME-38(12),** 1235–1245.

Enderle, J.D. (2000). The fast eye movement control system. In *The Biomedical Engineering Handbook Volume 2*, 2nd Ed. (J. Bronzino, Ed.). CRC, Boca Raton, FL, 166-1–166-2.

Enderle, J.D. (2002). Neural control of saccades. In *The Brain's Eyes: Neurobiological and Clinical Aspects to Oculomotor Research, Progress in Brain Research, V. 140* (J. Hyönä, D. Munoz, W. Heide and R. Radach, Eds.). Elsevier, Amsterdam, 21–50.

Fenn, W.O. and Marsh, B.S. (1935). Muscular force at different speeds of shortening. *J. Physiol. (London)* **35,** 277–297.

Hill, A.V. (1938). The heat of shortening and the dynamic constants of muscle. *Proc. R. Soc. London Ser. B* **126,** 136–195.

Hsu, F.K., Bahill, A.T. and Stark, L. (1976). Parametric sensitivity of a homeomorphic model for saccadic and vergence eye movements. *Computer Programs in Biomed.* **6,** 108–116.

Jacquez, J. (1996). *Compartmental Analysis in Biology and Medicine*. BioMedware, Ann Arbor, MI.

Keener, J. and Sneyd, J. (1998). *Mathematical Physiology*. Springer-Verlag, New York.

Khoo, M.C. (2000). *Physiological Control Systems: Analysis, Simulation, and Estimation*. IEEE Press, Piscataway, NJ.

Kuo, B.C. (1991). *Automatic Control Systems*. Prentice-Hall, Englewood Cliffs, NJ.

Leigh, R.J. and Zee, D.S. (1999). *The Neurology of Eye Movements*, 3rd Ed. Oxford Univ. Press, New York.

Levin, A. and Wyman, A. (1927). The viscous elastic properties of muscle. *Proc. Roy. Soc. (London)* **B101,** 218–243.

Robinson, D.A. (1964). The mechanics of human saccadic eye movement. *J. Physiol.* (London) 174, 245–264.

Robinson, D.A. (1975). Oculomotor control signals. In *Basic Mechanisms of Ocular Motility and their Clinical Implication* (G. Lennerstrand and P. Bach-y-Rita, Eds.). Pergamon, Oxford.

Robinson, D.A. (1981). Models of mechanics of eye movements. In *Models of Oculomotor Behavior and Control* (B. L. Zuber, Ed.). CRC, Boca Raton, FL.

Scudder, C.A. (1988). A new local feedback model of the saccadic burst generator. *J. Neurophys.* 59(4), 1454–1475.

Seidel, R.C. (1975, September). Transfer-function-parameter estimation from frequency response data-A FORTRAN program. *NASA Technical Memorandum NASA TM X-3286.*

Sparks, D.L. (1986). Translation of sensory signals into commands for control of saccadic eye movements: role of the primate superior colliculus. *Physiol. Rev.* 66, 118–171.

Van Gisbergen, J.A.M., Robinson, D.A. and Gielen, S. (1981). A quantitative analysis of generation of saccadic eye movements by burst neurons. *J. Neurophys.* 45, 417–442.

Westheimer, G. (1954). Mechanism of saccadic eye movements. *AMA Arch. Ophthalmol.* 52, 710–724.

Wilkie, D.R. (1968). *Muscle: Studies in Biology*, Vol. 11. Edward Arnold, London.

第 13 章　基因组学和生物信息学

Spencer Muse 博士

本章目录

本章主要内容：

- 与基因组学相关的分子生物学基本原理；
- 基因组项目所涉及的主要数据以及这些数据的采集方法；
- 基因组数据的实际应用；
- 生物信息学和 DNA 序列分析领域的主要研究方向；
- 主要的生物信息学数据库以及各种网络资源的使用。

13.1 绪论

2003年4月，人类基因组的所有30亿对碱基的全部测序工作宣告完成。作为现代科学的一个里程碑，这项成就为了解和治疗基因疾病带来了极大的希望，并且已经有充足的证据表明这些希望能够成为现实。1991年，当人类基因组计划（Human Genome Project，HGP）开始启动时，人类基因疾病中已知DNA序列的不到100种，而现在已经达到1631种。HGP的完成提前了将近3年，并节省了10%预算经费。而它所提供的数据量却超过了原来的预计量。该计划的成功取决于许多领域的创新成果，例如，有关DNA等化合物操作的各项基础分子生物学的突破；生产DNA测序仪的工程技术和制造技术的发展；机器人和实验室自动化技术的发展；用于解读测序所得数据的统计学方法的发展；为克服基因组学科学家们所面临的各种计算难题而开发的专用计算机软硬件系统等等。显然，人类基因组计划对于基础和应用层次的交叉学科研究都具有孵化作用。

人类基因组并不是基因组时代惟一的生物目标。截止2004年6月，科学家们共破译了1557种病毒、165种微生物和26种真核生物的全部基因组，其中包括疟原虫(Plasmodium falciparum)、酵母菌、水稻和人等。只有依靠技术上的不断进步才能加速基因组学的发展步伐，并减少数据获取的费用。此外，计算方法和统计学方法的发展对于海量数据的解读也是必不可少的。由于数据的积累速度已经超过了计算机处理器速度的增长速率，因此，更加突出了数学和计算科学的应用研究和基础理论研究的重要性。

本章将介绍有关基因组学实验数据采集的关键技术，以及数据分析的一些重要数学方法，并且也在适当的地方举例说明基因组学在医学上的应用。

13.1.1 中心法则：从DNA到RNA再到蛋白质

为了了解基因组技术，首先必须弄清楚如下3个重要概念：遗传信息是如何存储的；这些信息是如何处理的；遗传信息如何从上一代传递给下一代。

如图13.1所示，一般生物体的遗传信息都存储在脱氧核糖核酸（DNA）分子中。虽然有些病毒的基因信息存储在RNA中，但这只是一些例外，并不形成主流。如表13.1所示，通常用核苷酸（即碱基）对的数量来衡量基因组的大小，其数值的变化范围很大，有趣的是基因组的大小与生物体的复杂度之间并不是密切相关的。基因组的功能单位称为基因。本章下面叙述中的“基因”也是指一段连续的核苷酸，这种定义不太确切，一是因为基因的类型有很多；二是因为即使针对给定的某种基因，专家们也很难就基因所在的精确起始点和终止点达成共识。例如，图13.2所示的这类基因是结构基因，其编码用于指导蛋白质的形成。还有一类基因是RNA基因，它们不包含任何蛋白质信息，而是通过折叠成特异的三维结构与其他分子产生相互作用，从而在生物化学反应过程中起重要的作用。如图13.3所示，一般生命体

中常见的一种 RNA 基因就是转运 RNA（即 tRNA）基因。



图 13.1 DNA 分子由盘旋成双螺旋结构的两条核苷酸单链组成，碱基配对（A-T 和 C-G）所形成的化学键维持双螺旋结构的稳定

表 13.1 基因组的大小和组成（数据主要来源于 Taft 和 Mattic，2003）

生物体	基因组大小/($\times 10^6$ bp)	基因数/个
HIV 病毒（HIV1）	0.009	9
SARS 病毒（SARS）	0.03	14
支原体（M. genitalium）	0.58	470
纳古菌（N. equitans）	0.49	552
肝螺杆菌（H. hepaticus）	1.80	1875
大肠杆菌（E. coli）	4.64	4288
恶性疟原虫（P. falciparum）	22.85	5268
啤酒酵母（S. cerevisiae）	12.10	6000
果蝇（D. melanogaster）	122	13 600
线虫（C. elegans）	97	19049
拟南芥（A. thaliana）	115	25000
早期智人（H. sapiens）	3200	30000
家鼠（M. musculus）	2500	37000

通常所说的“基因”一般都指结构基因，因此，如无特别注明，本文以下叙述中的“基因”一词都指“结构基因”。生物体中基因的数量和种类是目前基因组学科学家们关注的重点。各种生物体所含有的基因数量各不相同，变化范围很大，微小的支原体只有 470 个基因，而巨大的植物体则有 60000 或更多个基因。寄生生物的基因数就更少了，如 HIV 病毒只有 9 个基因。据估计，人类基因组中的基因数量大约为 30000 个，这可能是人类基因组计划最惊人的发现，就在 1998 年科学家们还认为

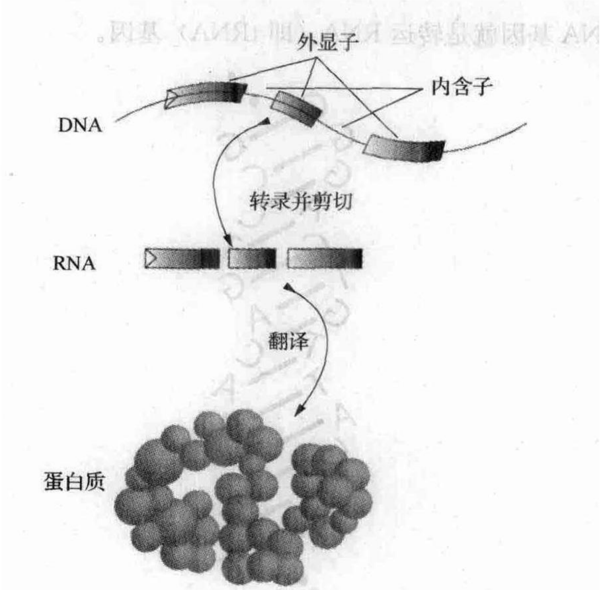


图 13.2　分子生物学中心法则

注：染色体上的基因 DNA 首先被转录形成包含内含子和外显子的 RNA，RNA 的内含子被剪除后，剩下的外显子连接起来形成成熟的 RNA 序列，后者经过翻译产生蛋白质

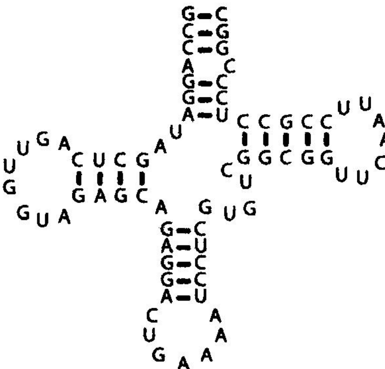


图 13.3　转运 RNA（即 tRNA）是一种典型的非蛋白质编码的基因

注：其功能取决于按照其 RNA 序列本身折叠形成的特异的二维或三维结构。

人类基因数量可能会达到 120000 个。这个估计值之所以有如此大的出入，其中一个原因是人类的基因组并不遵循许多真核生物基因组“一个基因一个蛋白质”的原则，相反，单个基因片断所包含的遗传信息可以翻译成多种具有不同功能的蛋白质。为了弄清楚这个问题，必须了解从 DNA 序列指令到功能性蛋白质形成这一过程所涉及的一系列步骤。如图 13.2 所示，分子生物学的中心法则表明：存储在 DNA 中的遗传信息首先被转录并形成 RNA，然后该 RNA 备份被翻译并产生具有特定功能的蛋

白质。

DNA 遗传信息复制到 RNA 备份中的这个过程被称为转录（transcription）（参见第 3 章）。很多人认为这个过程是从最早期的生命形式中遗留下来的，那时的生物体用 RNA 基因组存储遗传信息。转录过程中基因的表达受到调控，决定特定的基因在什么位置和什么时候开启或关闭。

转录开始时，RNA 聚合酶与基因的起始段结合，并延长产生出与染色体上 DNA 序列配对的信使 RNA 分子（mRNA）。这种 mRNA 分子将作为模板，翻译产生相应的蛋白质。不过，生物体必须调控基因的表达，以防止所有细胞在任何时候都将所有基因表达成蛋白质。这种调控由转录因子来实现，转录因子与染色体的 DNA 分子和 RNA 聚合酶相互作用，从而精确控制基因的表达过程。可以利用转录因子的反馈控制环来调控转录过程，如药物等外界环境的刺激引起转录因子的生成，从而触发基因的表达，而基因表达的产物反过来又会对转录因子产生作用。除了这种正向调控机制以外，还可以有负向调控。一种较新的学说认为，单个或者少数几个转录因子常常共同调控几套不同的基因，这几套基因通常共享一段上游 DNA 序列，而这段序列正是转录因子与基因的结合位点。

基因组时代最早的一个惊人发现是许多真核生物的基因序列并不相邻，而是被称为内含子的一些 DNA 序列片断隔开。如图 13.2 所示，在 RNA 被翻译成蛋白质之前，其实要先将 mRNA 序列上的内含子剪切掉。内含子的存在有助于解释生物体所产生的蛋白质数目比它所拥有的基因数目多的现象。内含子的剪切具有选择性，可变剪切使得外显子可以以不同的形式组合拼接，从而产生多种不同的蛋白质。例如，假设某个基因序列含有 E1、E2、E3 3 个外显子，分别被 I1 和 I2 两个内含子隔开。如果两个内含子都被剪切掉，那么将产生一个编码为 E1-E2-E3 的蛋白质。不过，也可能将外显子 E2 连同两个内含子一起剪切掉，从而得到 E1-E3 编码的蛋白质。就像调控基因表达的转录因子一样，也存在着调控选择性剪切的因子。这个领域的研究热点之一是寻找可以以多种不同方式剪切的单个基因，这种基因可以在特定组织中表达产生异构蛋白质。

按照 mRNA 分子模板的序列产生蛋白质的过程被称为翻译（translation）。理论上，翻译过程比转录过程和剪切过程要简单得多。开始时，核糖体与 mRNA 分子结合，然后核糖体以 3 个核苷酸为单位沿着 RNA 链移动。每 3 个核苷酸组合成一个密码子，编码 20 种氨基酸中的一种。核糖体每移动到一个新的密码子位置时，与 tRNA 产生相互作用并将该密码子所对应的氨基酸加到正在生长的蛋白质肽链上，然后核糖体继续沿 mRNA 向前移动到下一个密码子位置，如此重复进行（参见第 3 章）。

13.2 核心实验技术

新技术的发明实现了基因组科学各种数据的快速廉价采集，因而极大地推动了这门学科的迅速发展。这种发展一直被认为是数据驱动式的，而不是假说驱动式的，

研究人员首先收集大量的基因组数据，然后进行数据分析和相关实验，并获得对于遗传过程的进一步了解。虽然重要的生物技术日新月异，但是，如今以下 3 类数据是最重要的：基因组序列，它提供了我们认识生物遗传基础的起点；基因表达水平的检测数据，这些数据促进了基因调控的研究，与其他研究相结合，可以使我们了解生物体基因组与环境的相互作用过程；遗传的多样性，它使同一物种的个体之间存在差异，了解这些差异与疾病易感性等表型之间的相关性是现代生物医学研究极其重要的课题。

13.2.1 基因测序

虽然各种 DNA 基因的测序方法技术发展迅速，但是至今已经使用了几十年的基本原理却没有什么变化。使用最多的测序方法仍然是将“报道子”（reporter）附着在 DNA 核苷酸序列上，然后测量 DNA 序列在媒介中的迁移速度和距离。图 13.4 所示的桑格（Sanger）测序法就基于这个原理，它首创于 1974 年，简介如下。

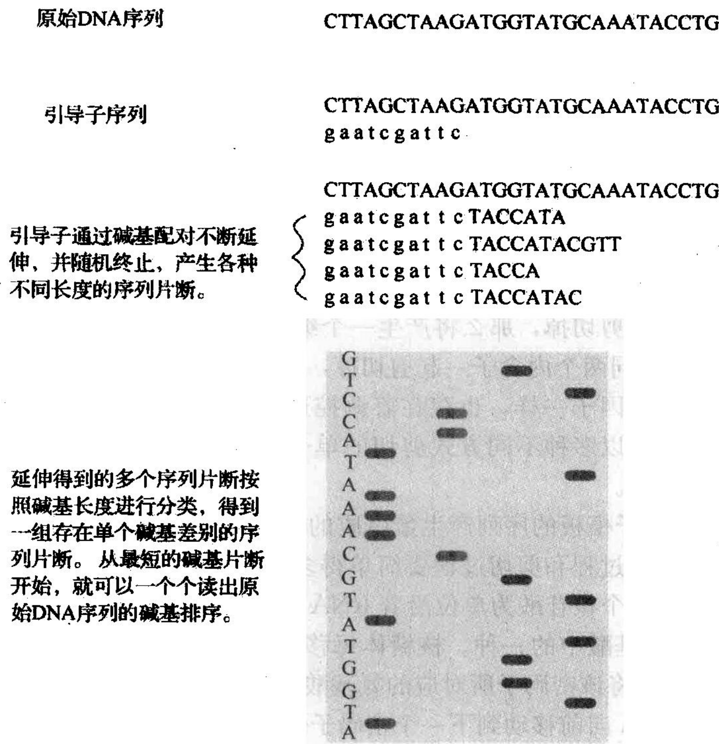


图 13.4 桑格 DNA 测序法的基本原理

DNA 序列具有方向性，通常将序列的 5′端作为左侧末端，3′端则作为右侧末端。桑格测序法首先从同一个 5′端核苷酸开始生成靶序列所有可能的序列，然后将带标记的报道子结合到每个序列片断 3′端最后那个核苷酸上。以前的报道子都用放射性

同位素标记，现在则用荧光进行标记。由于 4 种核苷酸各自对应于带不同荧光标记的报道子，根据荧光颜色就可以识别各个序列片段 3′端末尾的那个核苷酸。例如，假设待测序的 DNA 序列为 AGGT。以 5′端 A 开头的可能序列有 A、AG、AGG 和 AGGT，桑格测序法首先产生这 4 种序列片断，再在每个片断的 3′末端核苷酸上都加上报道子，然后根据这些序列片断在媒介中的迁移速率的不同，从短到长将序列分类。最短的序列就是 A，根据其报道子可知最后的核苷酸为 A；第二短的序列片断为 AG，其最后的核苷酸为 G，依此类推。将这些序列片断从短到长梯形排列之后，就可以直接根据每个序列片断的最后一个核苷酸读出整个 DNA 分子的序列了。

上述桑格测序法的基本方法已经经历了一系列的改进，大大提高了测序效率，可以用于测定包括人类基因组在内的整个基因组的序列。放射性报道子早已被安全且廉价的荧光染料所取代，并且，利用自动激光检测系统，在荧光标记核苷酸迁移的过程中就可以直接读出其碱基序列。早期的桑格测序法每次最多只能读出几百个核苷酸，如今先进的测序仪每次至少能够读出 800 个核苷酸。最重要的改进可能就是用毛细管电泳取代了平板凝胶电泳。老式的平板凝胶电泳工作量大且手工操作的重复性较差；新的毛细管系统实现了样品操作的自动化，使得测序过程几乎完全自动进行，不需要人工操作。在这一系列技术改进中，不可忽视的还有碱基自动识别算法的发明。其原理是：当核苷酸经过时，激光束测出 4 种报道子中每种的荧光发射强度，从而得到一张荧光色谱图；然后用标记软件“Pherd”等各种统计学算法，根据色谱数据计算得到具有极高准确度的 DNA 序列数据。这种自动识别算法大大减少了人工工作量。根据 4 条荧光强度曲线的相对水平，该算法不仅可以判别每个位点上可能性最大的核苷酸，而且还给出每个位点的误判概率。目前最新式的 DNA 测序仪每台每天可以检测多达 100 万个核苷酸。

13.2.2 全基因组测序

较长的 DNA 序列并不是整条进行测序，而是分成很多相互之间具有部分重叠的 DNA 短片断之后，再进行测序。如图 13.5 所示，最早由 Craig Venter 博士提出的鸟

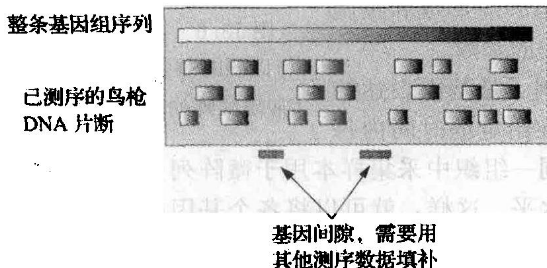


图 13.5　鸟枪测序法测定基因组序列或者其他大片断 DNA 序列的过程

注：首先将原基因长序列打断成许多 DNA 短片断，然后测定每个短片段的序列，最后识别基因片断的重叠末端，并将 DNA 片断进行排序组装，从而得到整条基因组序列。

枪测序法彻底改革了全基因组序列的测序方法，其基本原理很简单：建立大量基因组的相同备份；将基因组随机切割成几百万条 DNA 片断，再分别对每条短序列片断进行测序；通过鉴定片断末端配对的核苷酸将各片断重叠部分对齐，并依次排列起来；最后读出组装好的无间隙的 DNA 片断，从而得到整条基因组序列。在这种 Venter 方法提出之前，由于种种原因，鸟枪测序法一直被认为是不可行的，其中最让人望而却步的原因可能就是，对准百万条 DNA 片断的计算量太大了。为此，人们开发了专用的硬件系统以及相关算法来解决这个难题。

13.2.3　基因表达

高效率自动化测序技术实现之后，科学家们紧接着又开发了研究成千上万个基因产物的方法。这些方法其实就是高效率的新型 Northern 杂交法。对于基因样本中数以千计的基因，利用这些方法可以定量估计特定组织中每个基因所能产生的 mRNA 备份数。其中，cDNA 基因文库微阵列和寡核苷酸微阵列这两种方法目前在该领域占主导地位，它们为许多以往被认为不可能的研究工作打开了大门。下面举例说明其应用。

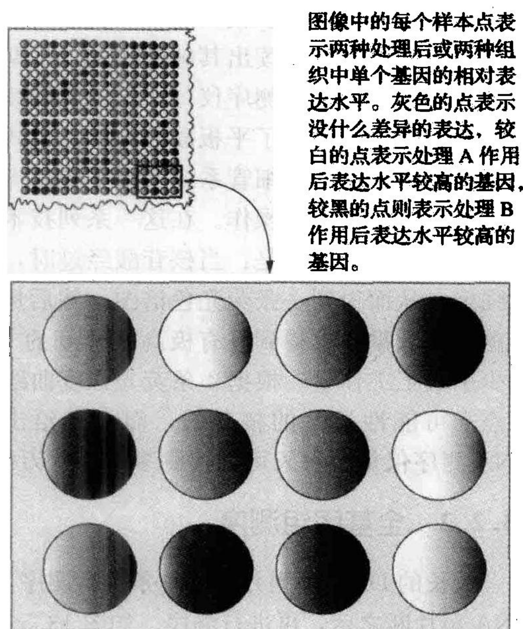


图 13.6　基因组学科学家应用微阵列技术可以同时测定数以千计的基因的相对表达水平（常规微阵列芯片至少可以包含 5000 个样本点）

例一，假设从某位癌症患者身上采集了两种细胞样本，一种是肿瘤细胞，另一种是正常组织细胞。应用微阵列分析法就可以识别出两种组织中同一套基因的不同表达水平。如图 13.6 所示，这套差异表达的基因中很有可能包含了许多与肿瘤形成和细胞增殖的生物过程相关的基因。

第二个常见的例子是基因表达的时间过程测定，即在给定的时间内每隔一定时间间隔从同一组织中采集样本用于微阵列分析。例如，用药之后，每隔 6h 检测一次基因表达水平。这样，就可以将各个基因的表达水平随时间的变化规律作图显示出来（见图 13.7）。如果阵列中某些样本点基因的表达水平一起增长或下降，那么可以判定这些点的基因为一群共调控（coregulated）基因，共调控基因都受到同一个转录因子的调控。

最后要介绍的微阵列技术应用的例子是其在医学诊断中的重要应用。假如医生能够得到病人肿瘤细胞的微阵列分析数据，那么，就可以将这些包含一批基因相对

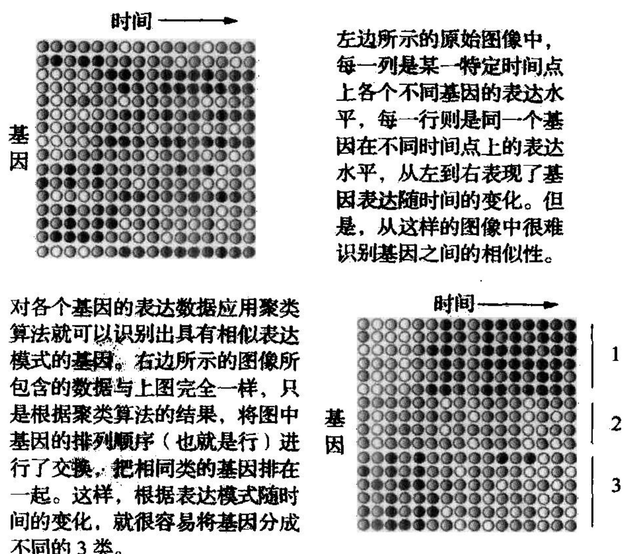


图 13.7　基因表达的时间过程可以用于研究基因表达随时间的变化

注：如果某些基因具有相似的表达模式，那么，就可以认为这些基因为共调控基因。

表达水平的数据与已知肿瘤类型的细胞中所采集的同样数据进行比较，如果病人的基因表达模式与某个参考样本的模式相符，就可以诊断该病人的肿瘤就属于参考样本的类型。随着微阵列技术的不断发展，各种癌症诊断的准确性迅速提高。

以上介绍了微阵列技术的一些应用实例。下面来看一下 cDNA 基因文库微阵列和寡核苷酸微阵列的基本原理。

cDNA 文库法是最早使用的微阵列法，并且至今为止仍然是最常用的方法，它是一种高产出率的基因表达检测法。操作过程如下：第一步，将数千种基因的 DNA 序列以点阵的方式附着到一张显微镜玻片上，每个样本点只含有单个基因的 DNA 序列。制作这种玻片的方法有很多，从简单的点样针装置到复杂的激光印制技术，层出不穷。第二步，采集靶细胞基因表达得到的所有 RNA，通过反转录过程，建立每个 RNA 的 cDNA。cDNA 分子是一条与某 DNA 链碱基序列互补的单链 DNA。例如，碱基序列为 GTTAC 的 cDNA 与基因序列为 CAATG 的 DNA 可以互补配对。构建 cDNA 文库时，用荧光染料将每个 cDNA 分子进行标记，然后将标记好的 cDNA 文库铺盖到整个玻片上，与已经附着在玻片上的 DNA 分子发生作用，待 cDNA 分子与相匹配的 DNA 分子结合后，洗去玻片上其余没有被结合的 cDNA 分子。第三步，通过测量微阵列中每个样本点的荧光强度，读出每个基因点所结合的分子数。基因的表达率越高，产生的 RNA 也越多，从而结合在点阵中的荧光标记 cDNA 也越多。

图 13.8 所示是实际应用中更为常见的基本 cDNA 法的一种改进方法。这种实验需要采集两种不同组织或个体的 RNA 样本，分别用两种不同的荧光染料标记其 cD-

NA，然后与同一块玻片上的 DNA 分子竞争杂交。根据两种荧光的相对强度就可以判断哪种 DNA 的表达水平较高。例如，可以测出某种特定基因在其中一种组织中的表达水平比另一种组织高 5 倍。

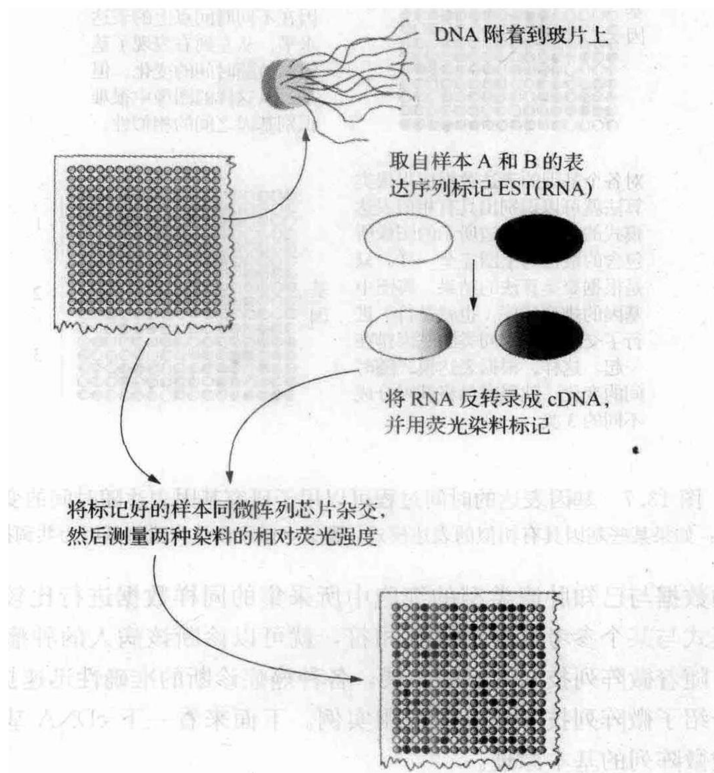


图 13.8　cDNA 微阵列芯片的制备与测定

注：步骤如下：1）将 DNA 以点阵的形式附着到玻片上；2）采集靶组织样本已经表达好的 RNA 序列，即表达序列标签（Expressed Sequence Tag，EST)；3）将 RNA 反转录成 cDNA，并进行荧光标记；4）将标记好的 DNA 分子与附着在玻片上的 DNA 进行杂交；5）通过测量荧光强度计算各个表达序列的量。

寡核苷酸微阵列分析法与上述方法略有不同，它分析的是 RNA 序列的相对表达水平。它不是将全基因组中每个基因的 DNA 序列都附着在玻片上，而是针对一些特定的基因的分析需要，采用短小的寡核苷酸链。也就是，为阵列中的每一个基因样本点设计 10～20 个长度约为 20～25 个碱基的不同寡核苷酸链，并将它们打印在芯片上。这种多寡核苷酸链的使用有利于减少各种干扰因素引起的误差。然后，将靶组织中采集的 RNA（而不是 DNA）进行荧光标记，并与寡核苷酸阵列进行杂交。这种方法的不足之处是每个芯片只能单独分析一个组织标本，无法进行竞争杂交。

虽然寡核苷酸微阵列法和 cDNA 微阵列法检测基因表达水平的基本原理相同，但是，它们具有各自的优缺点。上文已提到，寡核苷酸微阵列不能进行竞争杂交，并且其制备芯片的基因序列必须是现成的。因此，芯片制备费用很高，而且只能制

备已经商品化的或者典型生物体的寡核苷酸系统。与此相反，cDNA 微阵列的制备速度较快，即使是基因尚未测序的生物体也一样可以制备。不过，寡核苷酸微阵列法在给定的玻片上可以容纳较多的基因样本点，因此在同一张芯片上可以进行较多的检测，这也使得测量的重复性较好，这两个因素使得寡核苷酸微阵列的整体实验误差要比 cDNA 微阵列小得多，只是每次实验费用比较高。但是，如果将大量含误差的低价数据与少量高精度的高价数据进行比较，两者各有得失，并不能明确断定哪种方法具有明显的价格优势。

这两种方法都有一个缺点，就是只能测量 RNA 的水平，因此，它们只能是各种基因表达的蛋白质产物测量的暂时替代方法，直接测量蛋白质的方法更为理想且有用。虽然 RNA 水平与蛋白质水平具有相关性，但是这种相关关系成立的范围和程度还没有完全研究清楚。目前，由于高昂的费用，蛋白质微阵列系统的应用还很有限，不过，在不久的将来它肯定会成为主角。

13.2.4　多态性

人类基因组的“最终草图”公布于 2003 年 4 月，它包括了分布于 23 对染色体上的大约 29 亿个核苷酸组成的 30000～40000 个基因。人类基因组计划下一阶段的主要目标是揭示个体基因之间大大小小的各种差异是怎样导致表型水平的变化的。为了实现这个目标，科学家们正在努力探寻和整理有关基因多态性的数据。其实，多态性一直都是遗传学研究非常重要的课题。例如，人们对多线染色体结合模式的各种变化已进行了数 10 年的研究。20 世纪 60 年代曾经流行过同种异型酶的研究，就是根据氨基酸序列的整体电荷差异来进行分析的。不过，如今大多数遗传多态性的研究都集中于识别各个核苷酸水平的变异上。

国际单核苷酸多态性协会（网址为 http：//snp. cshl. org）是一个公立组织和私立组织的合作机构，该组织已经发现和表征了大约 180 万个人类单核苷酸多态性（Single Nucleotide Polymorphisms，SNP）。这对于医学具有非常重要的意义，这些单核苷酸个体变异的知识将加速人们认识基因疾病的进程，并且加快药物的开发。制药公司乐观地认为，有关个体变异的研究对于因人而异地选用药物具有重要的作用，并且可以根据病人各自的基本情况来分析药物可能带来的副作用。

大部分单核苷酸多态性是单个碱基被置换成另一个碱基引起的，少部分是插入或者缺失单个碱基。有关单核苷酸多态性的研究表明各个生物体内 SNP 的发生率互不相同，同一生物体不同部位的发生率也不相同。人类基因组中，大约每 1000～1500 个碱基中会有一个 SNP。不过，由于自然选择淘汰了种群中很多有害的等位基因，因此基因组中非编码区的 SNP 发生率要比编码区高得多。此外，同步多态性（即沉默多态性）的发生率比有义多态性或取代多态性要高，不过前者并不会使蛋白质编码发生变化。群体遗传学和分子进化论为有关 SNP 变异的研究提供了很多实验数据以及数学理论，可用于分析和预测各种生物状态下的 SNP 发生率。

简单重复序列（Simple Sequence Repeat，SSR）由长度为 2～13 个碱基的短小序

列串联重复而成，备份数一般为 10～50 个。SSR 的高突变率（mutation rate）导致 SSR 位点在多数种群中都是高变异性区域，因此 SSR 是一种非常重要的基因多态性来源。这种高变异性使得 SSR 成为个人身份鉴别的理想标记物，是 DNA 指纹识别的常用标记。人类的 SSR 位点通常含有至少 10 个等位基因，其每代的突变率约为千分之一。美国联邦调查局就利用一套含有 13 个四核苷酸的重复序列进行身份识别，有关专家声称，任意两个无亲属关系的人在所有这 13 个位点都含有完全相同的等位基因的可能性为零。

13.3 核心生物信息学技术

随着基因组数据采集技术的发展，对于管理和分析海量累积数据的新方法的需求也日益增加。“生物信息学”是应用数学、统计学和计算科学分析基因组数据的一门学科，其研究范围很广，包含数据库设计、系统工程、人工智能、应用数学和统计学等在基因组科学中的各种应用。

根据前面章节所述的有关基因序列分析的核心实验技术，现在我们可以来了解生物信息学的各个研究方向：1）为了将序列片断装配起来，必须进行序列比对（alignment)。2）所有这些序列包含了有关其来源、功能等方面的重要信息，这些信息都必须存储在数据库中，使得世界各地的用户都能随时获取。3）每次获得一段新的序列后，都必须对其进行功能预测。其中最基本的一项预测工作就是用计算机算法进行基因预测，即将基因组序列或染色体序列作为算法的输入，所得的输出就是预测的基因位置。4）无论是通过算法预测还是通过实验测定得到的基因序列，都必须对其进行功能预测，能够进行这项工作的生物信息学工具有很多。5）一旦有了微阵列数据之后，就必须对序列中的共调控基因群进行识别，并且鉴别出两种或两种以上不同处理情况下或者不同组织样本中差异表达的基因。6）利用 SNP 的多态性数据可以搜索相关性，例如，判断家族血统中是否存在某种疾病。这些研究方向都极其重要，它们涉及很多不同的学科领域，因此，生物信息学是一个高度学科交叉的研究领域。

13.3.1 基因组数据库

基因组工程是世界上不同领域的很多研究人员之间的一项长期合作项目，因此，必须使所有项目成员以及整个科学界都能够方便地获取有关研究结果所产生的数据。为了满足这种需求，研究人员已经建立了一些主要的数据库中心，用于基因序列的存储和读取，另外，有些数据库还提供高级序列分析方法和功能强大的可视化工具。

美国最主要的公共基因组学资源就是美国国家生物技术信息中心（National Center for Biotechnology Information，NCBI)，NCBI 的网址是 http：//www. ncbi. nih. gov，它提供了无数的基因组学数据和分析工具。在 NCBI 所收集的信息中，最重要的可能就是 GenBank 数据库，该数据库存储了 DNA 和 RNA 序列的信息。NCBI 还为用户提

供了各种 GenBank 搜索工具，并且在 GenBank 数据库的内容中设有链接，导向 NCBI 站内或站外的其他数据库。图 13.9 显示了 GenBank 核苷酸数据库的一个简略搜索结果。

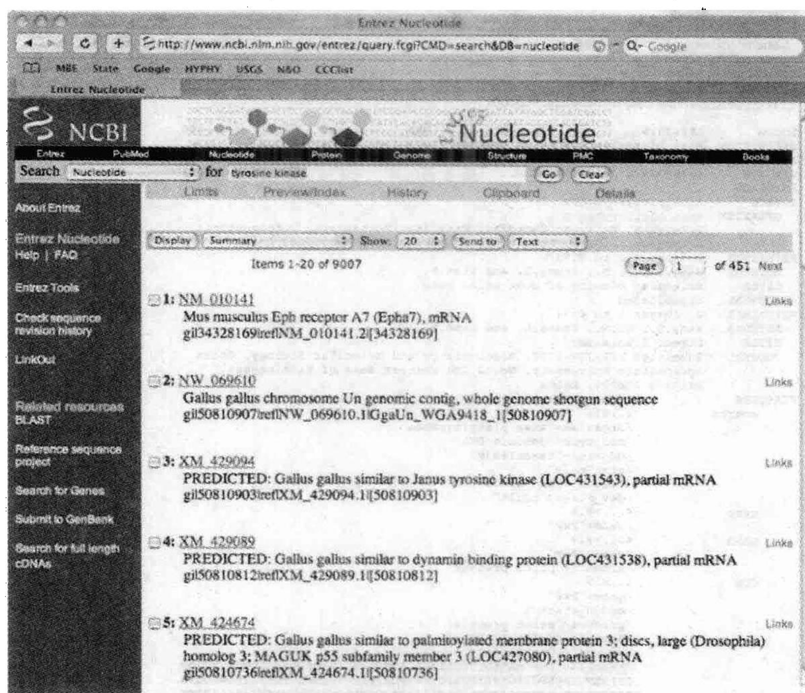


图 13.9　NCBI 网站上 GenBank 数据库的简略查询结果

注：此次检索在 GenBank 中共找到 9007 条包含“tyrosine kinase”的记录，单击各个记录条目可以查看其中的详细内容。

GenBank 数据文件所包含的内容非常丰富。图 13.10 显示了 GenBank 数据库中的鸭肮病毒蛋白的基因序列文件，编号 AF283319 是该条目特有的标识，文件中包含了长度为 819 个碱基的 DNA 序列、其预测的蛋白质序列以及测定该序列的中国实验室的相关信息。通过文件右上角的链接图标可以查询其他数据库的相关信息。熟悉 GenBank 数据库及其序列文件的各项内容是对基因组学和生物信息学工作者的基本要求。

此外，NCBI 还是数据库搜索工具 BLAST 的主页。BLAST 应用各种算法进行序列比对，在 GenBank 数据库中寻找与用户所查询的序列类似的序列。各种比对算法本章后面会介绍。为了说明 BLAST 的使用方法，下面来看一下哥本哈根大学 Eske Willerslev 教授的一项研究工作。Willerslev 教授和他的同事一起从西伯利亚永冻层采集了一批标本，其中含有约 30～40 万年之前保存下来的各种植物样本和动物样本。他们从这些标本的 rbcL 基因中提取出了多段短基因序列，然后将这些短基因序列输入 BLAST 算法，就得到了一系列相似序列的列表，其中最相似的几个序列很可能来

图 13.10 包含 DNA 序列的朊病毒蛋白基因的 GenBank 文件

源于古基因的生物体的近亲。

欧洲生物信息学协会（European Bioinformatics Institute，EBI）相当于欧洲的“NCBI”，其网址是 http：//www. ebi. ac. uk。浏览 EBI 网站之后，用户就会发现该网站与 NCBI 提供的功能几乎完全相同。特别值得注意的是“Ensembl 项目”（www. ensembl. org），它由 EBI 和 Sanger 研究所联合开发。Ensembl 拥有特别好的工具，可以通过其基因组浏览器（Genome Browser）搜寻基因组计划的有关数据。图 13.11 显示了人类 7 号染色体数据的一部分。在 Ensembl 中，可以与大鼠、小鼠、猩猩等已完成基因测序工作的其他动物的全基因序列进行比较，并且还可以进行基

因位置预测等。浏览器界面上所显示的条目一般都可以点击，以便察看更详细的内容。

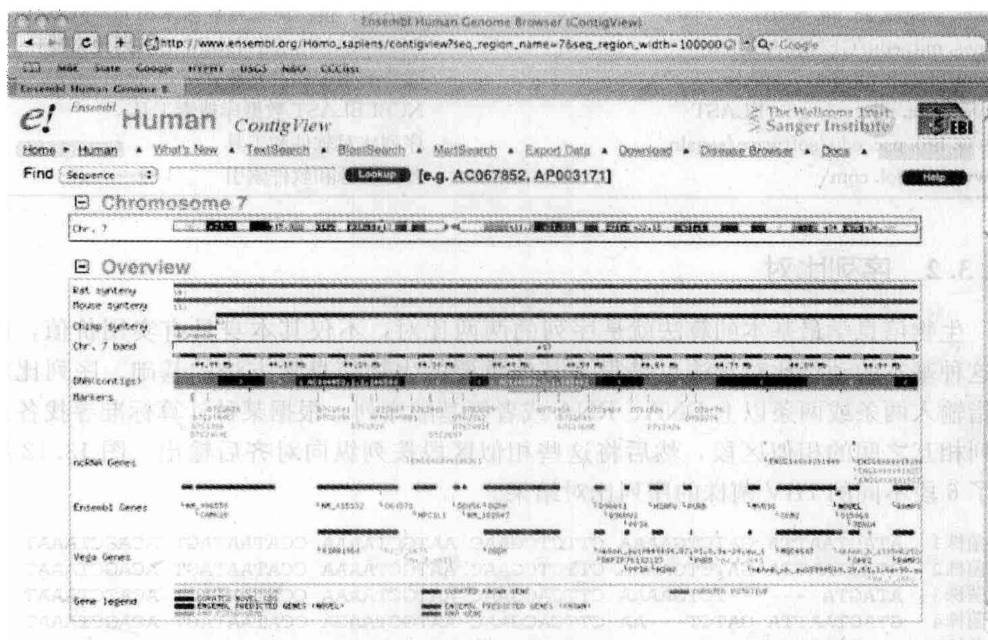


图 13.11　Ensembl 基因组浏览器中显示的人类第 7 号染色体的部分数据

基因组学科学家的日常工作中还用到很多其他重要的数据库和网络资源，表 13.2 列出了其中一部分，并简单描述了它们的特点。

表 13.2　部分在线基因组资源

综合性基因组资源	
www. ncbi. nlm. nih. gov	NCBI：GenBank 等
www. tigr. org	TIGR：微生物基因组，包含丰富的软件
www. ebi. ac. uk	欧洲生物信息学协会：数据库和分析工具
微生物基因组	
www. tigr. org/tigr-scripts/CMR2/CMRHomePage. spl	综合性微生物资源（CMR）主页
www. ncbi. nlm. nih. gov/genomes/MICROBES/Complete. html	NCBI 微生物基因组数据库
模式生物基因组数据库和资源	
www. arabidopsis. org	TAIR：拟南芥信息资源
www. wormbase. org	线虫数据库
flybase. bio. indiana. edu	FlyBase：果蝇基因组数据库
www. yeastgenome. org	酵母基因组数据库
genome. ucsc. edu	各种全基因组的浏览器

（续）

预测和分析工具资源	
www.sanger.ac.uk/Software/Pfam	PFAM：蛋白质家族数据库和 HMM 软件
genes.mit.edu/GENSCAN.html	GENSCAN：基于 HMM 的基因搜索器
www.ebi.ac.uk/arrayexpress	微阵列数据库及其软件
www.ncbi.nlm.nih.gov/BLAST	NCBI BLAST 数据库搜索工具
www.hto.usc.edu/software/seqaln	序列比对软件和工具
www.cellbiol.com	各种资源和软件索引

13.3.2　序列比对

生物信息学最基本的算法就是序列的两两比对，不仅其本身具有实用价值，而且这种基本的动态比对算法也是很多其他重要的生物信息学方法的基础。序列比对是指输入两条或两条以上 DNA、RNA 或者氨基酸序列，根据某种计算标准寻找各条序列相互之间的相似区段，然后将这些相似区段按列纵向对齐后输出。图 13.12 显示了 6 段不同的 HIV 菌株的序列比对结果。

```
菌株1  ATAGTAATTA GATCTGAAAA CTTCTCGAAC AATGCTAAAA CCATAATAGT ACAGCTAAAT
菌株2  GTAGTAATTA GATCTGAAAA CTTCTCGAAC AATGCTAAAA CCATAATAGT ACAGCTAAAT
菌株3  ATAGTA---- --TCTGAAAA CTTCACGAAC AATGCTAAAA CCATAATAGT ACATCTAAAT
菌株4  GTAGTAATTA GATCT---AA CTTCACGAAC AATGCTAAGA CCATAATAGT ACAGCTAAAT
菌株5  ATAGTAATTA GATCT---AA CTTCACGAAC AATGCTAAAA CCATAATAGT ACAGCTAAAG
菌株6  ATAGTAATCA GATCT---AA CTTCTCGGAC AATGCTAAAA CCATAATAGT ACAGCTAAAC
```

图 13.12　6 段 HIV 序列的多重序列比对

注：短线符号“—”表示原序列中插入或缺失的核苷酸位置。

序列比对有很多用途。序列的两两比对有助于判断两者是否具有相同或相似的功能。没有什么变异的比对区段很可能对应于蛋白质编码基因的重要结构或功能区域。通过分析几种物种基因比对的相似性情况，可以推测相关物种的进化历程，甚至还可以重建古生物的 DNA 或蛋白质序列。很多预测方法，包括蛋白质功能预测、转录因子结合位点的预测等，都要用到多重序列的比对结果。

为了说明序列比对的基本原理，下面来看一下两条 DNA 序列比对的例子。如果两条序列相似，那么它们很有可能来源于同一条很久以前的祖先序列。两段序列与祖先序列之间的差别以及两段序列之间的差别都是因为所经历的各种突变引起的。如图 13.13 所示，突变一般可以分成两类：一是碱基置换引起两段序列在突变位点产生差异（见图

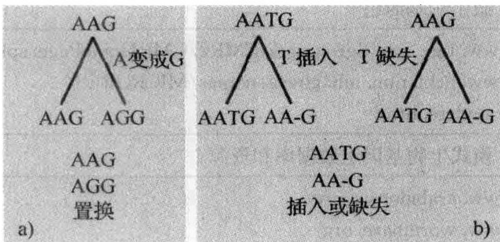


图 13.13　突变的两种类型

a）置换是单个核苷酸被替换成另一种核苷酸　b）插入或缺失（注意，插入和缺失的单个比对结果可能相同）

13.13a)；二是碱基插入或缺失（见图 13.13b）。在序列比对的示意图中常用短线符号“—”来表示碱基的插入或缺失，以下称之为“indel”。图 13.13b 显示了 indel 导致其中一段序列某些位点上存在的核苷酸在另一段序列中却没有。如果要准确无误的对齐两段序列，就必须了解这两段序列各个突变的全部历史过程。但是，这些信息一般不全，需要依靠计算机算法来重建各个突变事件可能发生的位点。比对算法的好坏可以用一个记分函数（score function）来衡量，各种算法的目标就是找出所有可能的比对中得分最高的比对。

例如，假设需要比对两小段残基 CAGG 和 CGA，那么，两序列对齐的可能方式共有 129 种，图 13.14 显示了其中的 8 种。如何判断这 129 种比对方式中哪一种最好

```
CAGG     CAGG-    CAGG     CAGG     CAGG-    CAGG     CA-GG    CAG-G
+ +      +  +     +          +      + +      +           +    + +
C-GA     C--GA    CGA-     -CGA     C-G-A    CG-A     --CGA    C-GA-
 a)       b)       c)       d)       e)       f)       g)       h)
```

图 13.14　序列 CAGG 和 CGA 总共 129 种可能的比对方式中的 8 种比对方式

呢？图 13.14 的 a 和 b 都有两个碱基匹配，不过，b 的比对有 3 个 indel，而 a 只有一个 indel；但是，a 有一对碱基不匹配，而 b 没有不匹配的碱基。其实，究竟哪个比对最好并没有绝对的答案，应该是匹配的碱基越多、不匹配和 indel 越少，比对就越好。我们可以用一种打分标准来量化这种直观的判断，给序列中的每一列比对碱基打一个分数 s_i，打分公式如下：

$$s_i=\begin{cases} m\text{，第 } i \text{ 列碱基匹配} \\ d\text{，第 } i \text{ 列碱基失配} \\ i\text{，第 } i \text{ 列有 indel} \end{cases}$$

如果设匹配分 $m=5$，不匹配分 $d=-1$，indel 分 $i=-3$，那么，图 13.14a 比对的得分为 $5-3+5-1=6$；同理，图 13.14b 比对的得分为 $5-3-3+5-3=1$。图中其他几个比对的得分分别为 0、0、1、0、-5 和 1。根据这种打分标准，可见比对 a 是最好的，实际上，它也是所有 129 种可能的比对方式中得分最高的比对。这个例子说明了寻找任意两个序列的比对中得分最高者的一种方法，也就是列举出所有可能的比对方式，计算每个比对的得分，再选出得分最高的那个比对。不幸的是，这种方法并不实用。可以证明，当两个序列的长度 n 较大时，其可能的比对方式大约有 $2^{2n}/\sqrt{2\pi n}$ 种，因此，即使只有 100 个碱基的一对短序列也有 6×10^{58} 种可能的比对方式，这个值比阿伏伽德罗常数还要大 35 个数量级呢！事实上，能有效识别最佳序列比对的方法，如 Needleman-Wunsch 算法和 Smith-Waterman 算法等，才是生物信息学重要的理论基础和切实可行的方法。

理论上，3 个或 3 个以上序列的比对与两两序列的比对方法相同，但是，计算量却随着参与比对序列数目的增加呈指数急剧增长，最终变得非常大。目前尚未找到切实可行的解决方法。这个问题属于一种基本计算难题，称为 NP 完全问题。除了计

算量的增加以外，两两序列的比对过渡到多重序列比对时，还出现了一个重要的新问题。多重序列比对时各列的打分不像两两比对时那样简单，由于各个序列的系统进化树不明确，很难制定出具有生物学意义的打分标准。不过，实际分析时通常不考虑这个问题，就直接按照两两比对的打分标准来求每一列中所有可能的两两配对的分数之和，将这个值作为多重序列比对中各列的打分值。例如，3 个序列比对中某一列的 3 个碱基是 CGG，如果用上述打分标准，即 $m=5$，$d=-1$，$i=-3$，则其总分为 $-1-1+5=3$。有一些算法如目前盛行的 CLUSTALW 软件则运用了一种渐进比对（progressive alignment）的方法来克服这个问题。

几乎所有常用的比对方法都基于与上述类似的打分标准来寻找最佳比对序列，显然，这种打分标准并没有什么生物学基础。而且，如何设定匹配、不匹配和 indel 的得分也是个问题。针对这些问题，人们已经做了大量的研究工作，并取得了一定的进展。最重要的改进就是用记分矩阵（scoring matrice）代替简单的匹配和不匹配记分，这种记分矩阵是根据氨基酸序列的实验数据建立的。例如，常用的 BLOSUM 记分矩阵和 PAM 记分矩阵中并不是给所有不匹配的列都赋予同样的 -1 分数，而是给每个可能的氨基酸配对赋予不同的罚分值，这些罚分值根据实验数据确定，化学特性相近的氨基酸之间的不匹配，如亮氨酸与异亮氨酸的不匹配等，所得的罚分要比化学特性差别较大的氨基酸少。

另一个改进是序列比对中 indel 罚分的设定。图 13.14b 和 e 中的两个比对虽然都有 3 个 indel，但是，b 中位点 2 和 3 的两个连续 indel 可能是由单次基因插入或缺失事件造成的。基于这种考虑，经常设置两种 indel 罚分，一是给第一个 indel 的开放罚分（open penalty），二是给后续 indel 的延伸 罚分（extension penalty）。例如，如果开放 罚分设为 $o=-5$，延伸 罚分设为 $e=-1$，那么 3 个连续 indel 的总罚分就是 $-5-1-1=-7$。

13.3.3 数据库检索

生物信息学中做得最多的工作就是检索 GenBank 等分子数据库，从中寻找与感兴趣的查询序列相似的序列。例如，查询序列是一段新近分离出来的流行病毒的基因序列，通过数据库检索就可以确定是否存在与这种新病毒相似的其他已知病毒序列。实质上，这种数据库检索就是一种特殊的双重序列比对，相当于把数据库中所有存在的序列都头尾相连串联起来，变成一条“超序列”，然后与查询序列进行比对。由于“超序列”的长度比查询序列要长很多很多倍，这种比对的输出结果中多数都是空隙，没有什么有用的信息。较为有效的处理方法是检查超序列中是否包含与查询序列匹配较好的小段序列，这种方法称为局部序列比对。Smith-Waterman 算法就可以确保找到局部比对中的最佳匹配。

虽然 Smith-Waterman 算法能够解决很多数据库检索问题，但是，对于大容量的 NCBI 等系统，由于每秒钟要处理很多条查询请求，该算法的速度还是太慢。为此，人们开发了各种启发式搜索工具，包括 BLAST 和 FASTA。虽然这些工具并不保证一定

能够找到最佳的局部比对，但是通常都能找到，因此它们都是很有用的研究工具。

可以毫不夸张地说 BLAST 是科学史上所有领域中最有影响力的研究工具之一，其网址是 http：//www. ncbi. nlm. gih. gov/BLAST。迄今为止已有 3 万多项研究引用了这个算法工具。BLAST 不仅提供了快速有效的数据库检索方法，而且也使与它相关的统计学理论快速发展，从而使得 BLAST 的应用迅速普及。如果要在一个非常大的数据库中搜索一小段序列，那么，在数据库中找到一条或者几条与查询序列相似的序列就是一种随机事件。因此，BLAST 将输出报表中的数据库匹配数据按照一个“E 值”进行分类，这个 E 值就等于找到具有某个匹配质量的匹配序列的随机概率。例如，E 值为 0.001，表示每一千次检索中平均只能找到一个这种匹配质量的匹配，表明该匹配具有重要的生物学意义；而 E 值为 2.0，则意味着平均每次搜索都可以找到两个具有这种匹配质量的匹配，因此这种匹配很可能没有什么意义。

例如，要在数据库中搜索 SARS 病毒的来源，其病毒蛋白的基因已被分离、测序并存入 GenBank 数据库，编号为 AY609081。如果将这段序列作为查询序列，在 NCBI 的 BLAST 中用 tblastx 进行搜索，得到的结果表明 GenBank 中最匹配的非 SARS 条目（注意，当时数据库中还未收录 SARS 条目）都属于冠状病毒（coronaviruse），这有力地证明了 SARS 是由一种冠状病毒引起的。这种基因组比较法已经成为流行病学研究的一个无价之宝。

13.3.4　隐马尔可夫模型

基因组学和生物信息学中的大量工作都集中在识别染色体或基因组等很长的 DNA 序列中的重要生物学功能区的研究上，基因、结合位点等很多重要区域在 DNA 上都表现为短小的不连续的邻近区域。隐马尔可夫模型（Hidden Markov Model，HMM）是特别适合这种特性识别的一种数学工具。HMM 在寻找河流的污染源、文字识别以及语音识别等各个研究领域都具有广泛的应用，因此，它具有成熟的理论基础。近几年，HMM 在解决基因组学新问题的应用研究中也取得了很大的成功。事实证明，在识别新基因组序列的基因位置、预测蛋白质的功能区、寻找内含子和外显子之间的边界以及预测蛋白质和 RNA 序列的高阶结构等方面，HMM 都是一种极其有用的工具。

为了初步了解 HMM 在 DNA 序列分析中的应用，我们来看一下等值区现象，等值区是指具有特定的碱基出现频率的各个 DNA 区域。人类基因组学的数据表明，由一百万或更多个碱基组成的 DNA 区域中，G 与 C 两个碱基的总含量为 20%～70%。如果 4 种碱基在整个基因组中的分布是均匀的，那么，该含量的变化范围不会这么大。如图 13.15 所示，按照一个简单的模型可以将基因组中每个碱基归属于 3 个不同的组：高 GC 组（H）、低 GC 组（L）和正常 GC 组（N）。正常组中的 A、G、C、T 4 种碱基的出现频率相等，都是 25%；高 GC 组中 4 种碱基的出现频率分别是 15% A、35%C、35%G 和 15%T；而低 GC 组中则为 30%A、20%C、20%G 和 30%T。在 HMM 算法中，这 3 个组称为“隐状态”，它们用碱基 A、C、G 和 T 的出现概率（emission probability）来描述。于是，利用这种简单的模型，基因组就可以看作是由

一系列分属于这 3 个隐状态的碱基区段组成（见图 13.15b）。

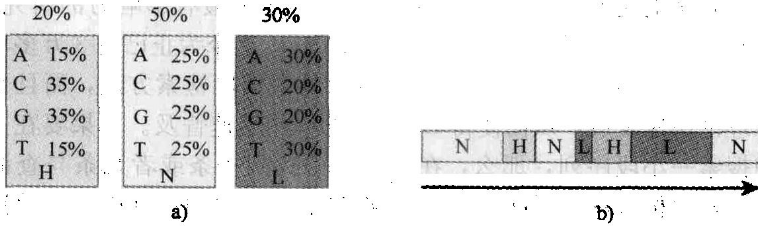


图 13.15 H、N 和 L 3 种隐状态及其碱基的出现概率（图 a），基因组中约有半数区段属于正常类 N，N 中每种碱基的出现概率均为 25%；基因组由属于 H、N 和 L 这 3 种隐状态的区段组成（图 b）。

本书不介绍 HMM 的详细数学定义，只简单介绍一下其基本概念。如图 13.16 所示，HMM 的每个隐状态都含有 4 种碱基，只是每种碱基的出现概率各不相同。模型还必须描述各个隐状态的模式和状态转换概率（transition probability），转换概率决定了单个隐状态区段可能的长度以及一个隐状态之后跟另一个隐状态的可能性。例如，一个低 GC 区之后是否可能跟一个高 GC 区，这就是由转换概率决定的。转换概率在基因预测等应用中具有非常重要的作用。

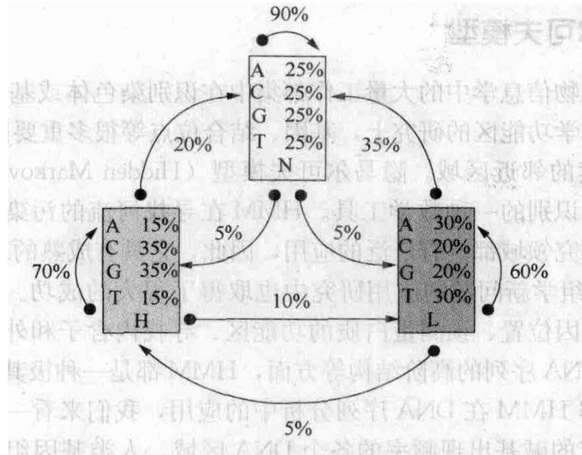


图 13.16 图 13.15 所示隐状态的一种 HMM 模型

注："转换概率"决定了一种隐状态紧随另一种隐状态的概率。例如，N 状态之后紧接另一个 N 状态的概率为 90%，紧接 L 状态的概率是 5%，紧接 H 状态的概率也是 5%。"出现概率"则决定了每个隐状态中 4 种碱基的出现概率。例如，隐状态 L 的 4 种碱基出现概率分别为 30%A、20%C、20%G 和 30%T。

了解了模型的基本组成之后，就可以应用现有的统计学方法，回答某个核苷酸链的有关问题，例如：

1）3 种隐状态的界线分别在哪里？

2）碱基 1200～1500 区段是属于高 GC 区吗？

3）碱基 3872 最有可能属于哪种隐状态？

4）高 GC 区段的长度小于 5000 个碱基的概率是多少？

下一节就举例说明这些问题的答案。

13.3.5　基因预测

基因预测的概念说起来很简单，就是给定一段 DNA 长序列，确定其中的基因位置。然而，这项工作其实并不那么简单。最容易想到的方法，可能就是找出序列中所有成对的起始密码子（ATG）和终止子（TAG、TGA、TAA）。如果起始密码子和终止子之间有较长的序列片断（如 300 个碱基）且其长度为 3 的整数倍，那么，这些序列很有可能就是蛋白质编码基因。虽然用这种简单方法可以找到很多基因，但是它具有很高的假阳性率（即将非基因序列错误地判定为基因）和假阴性率（即没有检出真正的基因）。并且，这种方法没有考虑内含子存在的可能性，无法预测短基因。切实可行的基因预测算法在预测基因时还要用到许多其他信息，包括基因的已知结构、已知基因的内含子和外显子的长度分布以及已知调控序列的共有序列。

HMM 很适合用于基因发现算法，图 13.17 显示了一种简单的 HMM 基因预测算法的基本结构。注意，该 HMM 模型包含有多个隐状态，即启动子区域、起始密码子、终止密码子、外显子、内含子以及落在不同基因之间的非编码区。另外，并不是每个隐状态之间都存在连接，而是根据基因和基因组结构的信息设定连接。例如，“起始密码子—内含子—终止子—外显子—启动子”这样的排序不存在生物学可能性，而“非编码区—启动子—起始密码子—外显子—内含子—外显子—内含子—外显子—终止子—非编码区”这样的排序是可能的。好的 HMM 模型必须广泛收集并结合此类生物学信息。

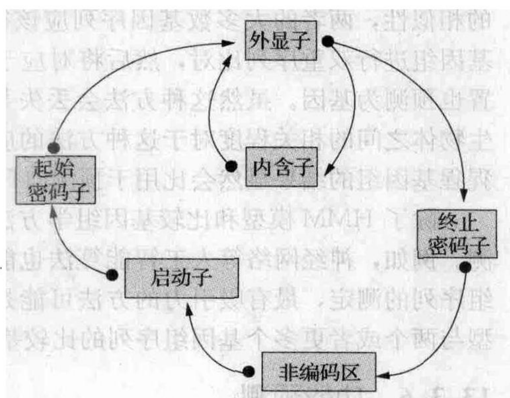


图 13.17　一种简单的基因预测 HMM 模型

图 13.17 的 HMM 算法在投入使用之前，必须先进行训练。训练内容中包括利用已知基因现成的序列估算模型各个隐状态的转换概率和出现概率。例如，如果训练数据中含有 150 个内含子，那么，这些内含子中测得的碱基 C 的频率就可以作为模型中内含子隐状态 C 的出现概率。内含子和外显子的平均长度则可以用来估算内含子和外显子隐状态的转换概率。训练完成后，这个 HMM 算法就可以用来预测长 DNA 序列上的基因位置、内含子和外显子的边界以及启动子位点等，以此类推。

实际使用的基因预测算法要比图 13.17 所示的模型复杂得多，但是基本结构是相同的。经过越来越多基因组的测算，基因预测算法的性能不断提高，HMM 模型也不断得到改进。对于细菌基因组，新型的基因预测算法已经很少出错，有高达 95%

的预测基因后来被证实是真基因，只有1%～2%的真基因被算法遗漏。然而，对于真核生物的基因预测，情况却没有这么乐观，真核生物的基因组要大得多，并且基因的结构也复杂得多，最值得注意的是真核生物的基因包含有内含子。基因预测算法的性能通常用灵敏度（sensitivity）和特异性（specificity）来衡量。如果这两个参数都用单个碱基为单位来计算，那么，灵敏度 S_n 定义为预测到的真基因碱基数占实际真基因碱基总数的百分比；特异性 S_p 则定义为预测的基因碱基数中确实存在于实际真基因中的碱基所占的百分比。好的基因预测方法具有高灵敏度和高特异性。如今最好的真核生物基因预测算法各个碱基的灵敏度和特异性都可以达到90%左右。如果这两个参数以整个外显子为单位来计算，也就是统计算法预测整个外显子位置是否正确，那么，灵敏度和特异性的值就会下降到70%左右。

基因位置预测的一种新方法是利用比较基因组学（comparative genomics）的方法，这种方法功能很强。目前科学家们已经获得了人类全基因组序列，也探明了上万个基因的位置。假设某个实验室现在正在测定猎豹（cheetah）的基因组序列并预测基因的位置。那么，由于人和豹都属于哺乳动物，两者的基因组中必定存在一定的相似性，两者的大多数基因序列应该很相似。因此，基因预测时，可以先对两个基因组进行双重序列比对，然后将对应于已知的人类基因的那些猎豹基因组上的位置也预测为基因。虽然这种方法会丢失某种物种特有的基因，但它非常有效。两种生物体之间的相关程度对于这种方法的应用有很大的影响，人类基因组用于预测大猩猩基因组的结果显然会比用于预测向日葵和草履虫基因组要好得多。

除了HMM模型和比较基因组学方法以外，还有其他很多方法可以用于基因预测。例如，神经网络等人工智能算法也能有效地进行基因预测。随着越来越多基因组序列的测定，最有吸引力的方法可能是各种算法的综合，例如，可以将HMM模型与两个或者更多个基因组序列的比较基因组学数据相结合。

13.3.6 功能预测

随着基因组序列的测定，以及序列上的基因预测的实现，接下来需要完成的生物信息学工作就是确定基因的生物学功能。理论上，可以通过分子生物学的实验研究来确定每个基因的功能，但是，实际上这种方法显然是不可行。下面介绍功能预测的两种基本计算机算法，一种利用比较基因组学，另一个则利用HMM模型。

比较基因组学方法用于基因功能预测的依据是一个简单的逻辑推理，即如果物种A的某个基因与物种B的某个基因很相似，那么，这两个基因很可能具有相同或相关的功能。其实，这个推理一直用于较高层次的生物学研究，例如，尽管物种之间存在着许多区别，但不同动物的肾脏都具有相同的基本生物功能。在基因层次上，这种推理不那么准确，尤其当被比较的物种之间亲缘关系不很密切时，这种推测就更不准。但是，无论如何这种方法还是很有用，并且通常很有效。

简单的数据库搜索是利用比较基因组学预测基因功能的最直接方法。将新发现的基因序列输入到BLAST中之后，如果它与细胞色素氧化酶的基因相匹配，那么，

这个基因本身就很可能是细胞色素氧化酶基因。但是，如果新基因与远亲物种的基因序列相匹配，或者匹配区段太短，或者新基因序列与某个多基因家族的多个成员都能匹配，那么，情况就比较复杂了。下面详细分析这 3 种情况。在第一种情况下，两个物种数千万年之前的共同祖先经历这么多年的演变之后，各自的基因功能可能早已发生变化。但是，如果两个或更多个这样的远亲物种具有几乎完全一致的基因序列，那么，就很有理由说明这个基因对于两种生物体都非常重要，并且在漫长的进化历程中一直保持着同样的生物功能。第二种情况的短匹配片断可能只是由蛋白质的基本结构相同引起的。例如，由于都存在编码 α 螺旋的基因片断，两个基因序列就可能有部分区段相匹配。这种匹配提供了很有用的结构信息，但是没有较强的证据可以说明两者具有相同的功能。第三种情况的多基因家族是由基因复制以及随后产生的功能分化引起的，典型的例子有球蛋白基因家族和淀粉酶基因家族。基因家族的产生过程简述如下：在某个时期，生物体基因组的某个单基因被完全复制，并且复制的基因备份当时可以自由演变成具有新功能的基因，这种新功能通常与以前的功能具有相关性。后来的一次次复制就导致了这些多基因家族的生长和分化。由于基因家族的所有成员都拥有共同的祖先基因，因此，它们可能含有相似的 DNA 序列。如果在数据库搜索中，新基因序列与基因家族的多个成员都有匹配，那么，就很难精确判断新基因的功能，不过，通常可以限定在一类功能之内。

利用比较基因组学的方法，所有已知的蛋白质都已经按功能被分类。假如在 GenBank 蛋白质数据库中查询蛋白质序列 A，得到的最佳匹配是蛋白质序列 B；反过来，再在数据库中查询蛋白质 B，如果所得的最佳匹配序列是 A，那么，可以认为这两个蛋白质相互之间最相配，并且两者很可能具有相同的功能。同理，如果序列 A 的最佳匹配是序列 B，B 的最佳匹配是序列 C，而 C 的最佳匹配又回到 A，那么，序列 A、B、C 很可能具有相同的功能。利用这一原理，可以将预测功能相似或者相同的基因进行聚类，NCBI 网站上的蛋白质直系同源簇（Clusters of Orthologous Groups，COG）数据库就是用这种方式建立起来的，它是整个 GenBank 蛋白质数据库内所有蛋白质的综合性聚类，其网址是 http：//www. ncbi. nlm. nih. gov/COG。

在已知的蛋白质和各种蛋白质结构域中，具有相同功能或相同结构的例子有很多。蛋白质家族数据库（Protein Family，PFAM）就包含了近 7500 个这样的蛋白质家族的多重序列比对，其网址是 http：//sanger. ac. uk/Software/Pfam。PFAM 项目组人员利用每个比对的序列数据建立了一种特殊的 HMM 算法，称为模式 HMM。PFAM 数据库可以接受查询序列，利用这 7500 个蛋白质家族的数据及其关联的模式 HMM 算法，从而判断该序列是否属于某个基因家族。PFAM 查询的输出结果中包含了查询序列分别属于各个蛋白质家族的概率，不仅指出了最佳匹配，而且还标出了匹配的强度。目前，GenBank 数据库中大约有 74%的蛋白质序列与 PFAM 数据库中的蛋白质家属相匹配，这表明每个新发现的蛋白质在 PFAM 数据库中找到匹配序列的可能性相当高。PFAM 备受关注的原因不仅因为它很有效，还由于它具有比较基因组学与 HMM 相结合的理论方法。

13.3.7 差异表达基因的识别

利用微阵列和寡核苷酸阵列技术测定经过两种不同“处理”的几千个基因的表达水平，这是一个常规实验。通常其中一种处理是对照处理；另一种则是外界刺激，如药物、化学药品或者温度和 pH 等物理量的变化。也可能是两种不同器官组织之间的比较，如大脑组织与心脏组织；或者患病组织和健康组织之间的比较，如肿瘤组织与正常组织；或者同一组织在两个不同发育阶段的样本之间的比较，如胚胎与成人。进行此类实验的主要目的之一是鉴别两种不同“处理”作用后产生表达差异的基因。

假设某个简单的双处理微阵列实验的基本数据如下：

样 本 点	处理 1	处理 2	比 率
1	350	250	1.4
2	1200	1250	0.96
3	900	300	3.00
4	14000	52 200	0.27

表中，样本点指微阵列上的位置；处理 1、处理 2 为经过两种处理的基因表达水平的原始测定数据；比率是处理 1 测量值/处理 2 测量值。

微阵列上每个样本点对应于一个基因，在竞争杂交实验中，要从每个样本点测得两种不同“处理”所产生的基因表达情况。注意，上述表格中第一栏的名称故意用“样本点”，而不是“基因”。这是由于每个基因需要多次重复测量，因此总有不同的样本点加的是同一个基因。表中最后一栏是这个实验最重要的数据。由表中可见，样本点 4 的基因相对表达的差异最大，“处理 2”作用下的基因表达水平几乎是“处理 1”的基因表达水平的 4 倍。现在的问题是，比率达到多大或者多小时才能断定基因表达水平确实有差别呢？这是一个变异性和统计显著性的问题。换言之，如果重复进行样本点 4 的实验，那么，所得到的比率可能都接近 0.27 吗？上述表格中的数据不足以回答这个问题，这说明了重复实验的重要性。任何定量测量数据之间的比较都需要重复实验，以便估计测量的方差。早期微阵列数据分析中，人们往往忽视这个实验设计的基本原则。不过，现在的研究工作都非常注意实验设计并应用方差分析（ANOVA）合理地处理数据。如今常规的实验对每个基因都要进行至少 5 次重复测量，如果要检测基因表达水平很细微的处理差别，就需要更大量的重复测量。

13.3.8 相同表达模式基因的聚类

微阵列实验的另一个目标不是检测差异表达的基因，而是要鉴别两种或两种以上处理的作用下表达模式相似的基因群。这种研究工作最好使用基因表达的时间过程数据，也就是在基因表达的过程中，按照一定的时间间隔测量一系列基因表达水平的数据。例如，实验小鼠接触有毒化学物质之后，每隔 1h 测量一次其基因表达水平；妊娠期每隔 3 个月检测一次母亲或胎儿的基因表达水平；每隔 1 年检测一次

HIV 感染病人的基因表达水平，等等。如果把时间作为横坐标 x 轴，把基因表达水平作为纵坐标 y 轴，那么得到的各种基因表达水平的叠合图如图 13.18 所示，通过目测就能比较基因的表达模式，按照表达水平增减变化的一致性将基因进行分类。由图 13.18 可见，基因 1 与 4 的表达模式很相似，基因 2 与 5 也很相似。

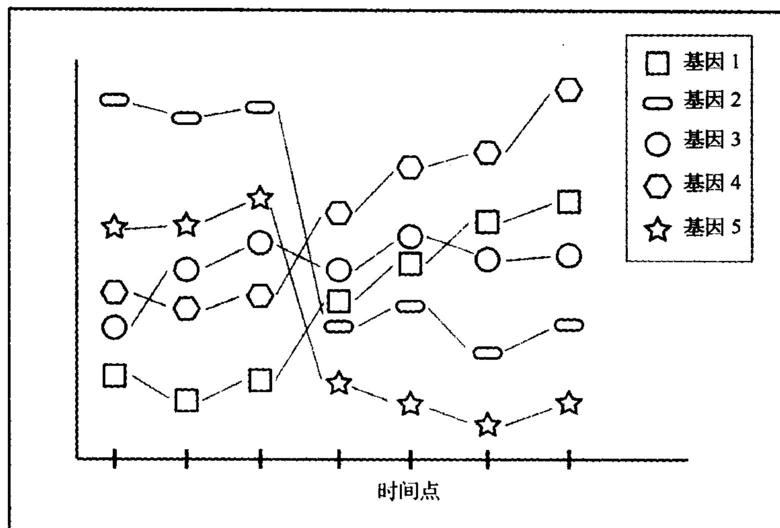


图 13.18　5 个基因表达模式的叠合图

注：基因 1 与 4 以及基因 2 与 5 之间具有相似的表达模式，表明这两对基因可能分别受到相同的转录因子控制。

两种基因表达模式之间的相似程度可以用相关系数（correlation coeffcient）来描述，即

$$r_{X,Y}=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{S_X}\right)\left(\frac{Y_i-\overline{Y}}{S_Y}\right)$$

式中　X_i、Y_i——基因 X 和 Y 在时间点 i 上的表达水平。

如果相关系数的值接近 1 或－1，表明两种基因具有非常相似的表达模式。但是，面对成千上万的表达模式数据，计算相关系数的工作就比较麻烦了。根据基因表达模式的相似程度进行基因分类是一个研究热点，这方面的算法已经发表了很多。所有这些基因聚类算法的目标都是要使同一类基因表达模式的差异尽可能小，而不同类之间则具有一定的差异。“从上至下”聚类算法开始时把所有基因都看成一个大类，然后反复将各个基因分到越来越小的不同的类中去。相反，“由下至上”聚类算法开始时将每个基因都看成一类，然后逐渐将小的类合并，产生越来越大的类。聚类算法可以有人工干预，首先设定最终分类的数目，也可以自动运行，由算法决定最终的分类数目。

13.4　结论

基因组学这门学科的诞生不仅为分子生物学的研究提供了丰富的工具和数据和

研究手段，而且极大地促进了众多交叉学科研究的惊人发展，从而对高等教育机构长期以来所维持的体系设置提出了挑战。新一代生物学家操作计算机工作站时应该与操作实验台一样自如。基于这种共识，很多大学已经重新调整了科系和课程设置，以适应基因组学发展的需要。

在实际应用中，基因组学的研究成果将会逐渐进入医学应用，实际上，随着基因组学的发展，医学诊断方法正在迅速产生变化。基因组学的下一阶段研究重点将放在建立基因型与复杂的表型之间的关系上，随着这些关系的建立，新的治疗方法和药物就会应运而生。例如，如果有一种药物对 99%的使用者具有显著的疗效，但对剩余的 1%使用者却会产生严重的副作用。这样的药品现在很难在市场上出售。但是，如果利用基因筛选的方法区分出可能产生副作用的患者，那么，这些药物就能够安全有效地使用了。另外，尽管还有些遥远，但将来基因疗法肯定可以治疗基因缺陷导致的疾病。生物学、工程学和数学等学科的不断交叉合作将推动这些前沿研究领域的发展。

习　　题

1. 含有 4 个外显子的一段基因片断可能产生几种蛋白质？

2. 一般来说，真核生物有内含子，而原核生物没有。请分析内含子的优缺点。

3. 多数氨基酸可被一个以上密码子编码。如果生物体多个同义密码子中的一个使用率特别高，对该生物体的基因组的组成可能会造成什么影响？如何应用这种现象设计寻找基因的算法？

4. 一段长为 20 个碱基的寡核苷酸与任意一段非设计匹配序列的匹配概率是多少？为了简化，假设 4 种碱基的出现概率均为 25%。人类基因组中能与这段寡核苷酸匹配的序列可能有多少？

5. 如果 FBI 用于身份识别的一套 13 个简单重复序列标记中的每个标记都含有 20 个频率相同的等位基因，那么，任意两个人的这 13 个标记都含有相同等位基因的概率是多少？

6. 有多少哺乳动物的基因组序列已经完成测定？是哪些哺乳动物？

7. 疟蚊（Anopheles gambiae）基因组的大小是多少？它有几条染色体，含有几个基因？

8. 果蝇乙醇脱氢酶的基因长度是多少？

9. 如下是序列片断 CGGTCA 和 CAGCA 的两个序列比对：

```
C-GGTCA        C-GGTCA
CA-G-CA        CA-GCA-
```

(a) 如果匹配得分为 5，不匹配罚分为 −2，空位罚分均为 −4，请计算两个比对各自的分值。

(b) 如果开放空位罚分为 −5，延伸空位罚分为 −1，请重新计算两个比对各自

的分值。

10. 假设某台计算机每秒可以计算 100 万个比对的分值，那么，需要多少时间才能完成长为 1000 个碱基的两段序列的穷尽搜索以便找出最佳比对?

11. 试利用 GenBank 查找一段锌指基因序列，并使用 BLAST 搜索工具检索 GenBank 中有多少序列与你所找到的序列相似，你能从搜索结果中获得有关锌指基因的什么信息?

12. 图 13.17 所示简单的基因发现 HMM 算法中还可以添加哪些其他特性?

13. 请作图表示适用于原核生物的简单的基因发现 HMM 算法，该算法应该包含外显子和基因间片断的隐状态，并且必须确保每个外显子的长度为 3 的整数倍。

14. 请利用 PFAM 网站的资源，简单描述结节性硬化基因家族成员的结构和功能。

15. 图 13.18 中的基因 2 的表达水平比基因 5 要高，请证明这 2 个基因具有相似的表达模式，它们可能是共调控基因。

16. 4 个不同时间下测得的 3 个基因的表达水平分别如下：

基因 1	5	7	20	22
基因 2	15	12	18	17
基因 3	10	15	50	45

请计算两两基因表达模式的相关系数，并确定是否存在相似的模式。

17. 4 个不同时间下测得的 2 个基因的表达水平分别如下：

基因 1	12	16	24	28
基因 2	64	16	16	64

请计算这对基因表达模式的相关系数，并以基因 1 表达水平为纵坐标，基因 2 表达水平为横坐标，作图表示两者之间的关系。图中所示的情况是否与相关系数所表明的含义一致? 请基于你所观察到的结果说明用相关系数可能产生的错误。

18. 微阵列芯片上的基因序列的功能常常是未知的。假设有一个实验要鉴定一个基因，该基因对某种特定肿瘤的形成非常重要。请解释如何利用 NCBI 的资源来确认这个基因的功能。

19. 图 13.7 所示的基因表达的时间过程研究中，需要根据表达模式进行聚类，请设计一种基因聚类的算法。

20. 在使用 BLAST 工具检索一段蛋白质编码基因时，可以用两种方法：一是用原始 DNA 序列作为查询序列；二是将原始 DNA 序列翻译成氨基酸序列进行查询。请说明这两种方法中哪种更好，两者的优缺点各是什么。

推荐阅读资料

Altschul, S., Gish, W., Miller, W., Myers, E. and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Bernardi, G. (2000). Isochores and the evolutionary genomics of vertebrates. *Gene* **241**, 3–17.

Brown, P.O. and Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. *Nat. Genetics* **21 (suppl)**, 33–37.

Burge, C. and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94.

Burris, J., Cook-Deegan, R. and Alberts, B. (1998). The Human Genome Project after a decade: Policy issues. *Nat. Genetics* **20**, 333–335.

Cantor, C.R. and Smith, C.L. (1999). *Genomics: The Science and Technology Behind the Human Genome Project*. John Wiley and Sons, New York, NY.

Collins, F.S. et al. (1998). New goals for the US Human Genome Project: 1998–2003. *Science* **262**, 682–689.

Fraser, C.M. et al. (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397–403.

Gibson, G. and Muse, S.V. (2002). *A Primer of Genome Science*. Sinauer Associates, Sunderland, MA.

Hartl, D. and Clark, A. (1998). *Principles of Population Genetics*, 3rd Ed. Sinauer Associates, Sunderland, MA.

Henikoff, S. and Henikoff, J. (1992). Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* **89**, 10915–10919.

International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.

International SNP Map Working Group. (2001). A map of human genome sequence variation containing 1.42 million single-nucleotide polymorphisms. *Nature* **409**, 928–933.

Kerr, M.K., Martin, M. and Churchill, G. (2000). Analysis of variance for gene expression in microarray data. *J. Comput. Biol.* **7**, 819–837.

Lee, C., Weindruch, R. and Prolla, T. (2000). Gene-expression profile of the aging brain in mice. *Nat. Genetics* **25**, 294–297.

Mount, D.W. (2001). *Bioinformatics: Sequence and genome analysis*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Needleman, S. and Wunsch, C. (1970). A general method applicable to the search for similarities in the amino acid sequences of two proteins. *J. Mol. Biol.* **48**, 443–453.

Scherf, U. et al. (2000). A gene expression database for the molecular pharmacology of cancer. *Nat. Genetics* **24**, 236–244.

Smith, T. and Waterman, M. (1981). Identification of common molecular subsequences. *J. Mol. Biol.* **147**, 195–197.

Sonnhammer, E., Eddy, S., Birney, E., Bateman, A. and Durbin, R. (1998). Pfam: Multiple sequence alignments and HMM profiles of protein domains. *Nucl. Acids Res.* **26**, 320–322.

Taft, R.J. and Mattick, J.S. (2003). Increasing biological complexity is positively correlated with the relative genome-wide expansion of non-protein-coding DNA sequences. *Genome Biology* **5**, P1.

Venter, J.C., Adams, M., Sutton, G., Kerlavage, A., Smith, H. and Hunkapiller, M. (1998). Shotgun sequencing of the human genome. *Science* **280**, 1540–1542.

Venter, J.C. et al. (2001). The sequence of the human genome. *Science* **291**, 1304–1351.

Waterman, M.S. (1995). *Introduction to Computational Biology*. Chapman and Hall, London.

Wheeler, D.L. et al. (2000). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **28**, 10–14.

Willerslev, E. et al. (2003). Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* **300**, 791–795.

Zhang, L. et al. (1997). Gene expression profiles in normal and cancer cells. *Science* **276**, 1268–1272.

第 14 章 计算生物学与复杂性理论

Charles Coward 医学博士⊖
Banu Onaral 博士⊖

本章目录

本章主要内容:

- 计算机模型如何用于细胞生理过程的建模仿真;
- 细胞调控的简化模型举例;
- 复杂性理论的基本概念;
- 复杂性理论在细胞生物学中的应用。

14.1 计算生物学

计算生物学的研究内容包括新陈代谢和调控途径等各种细胞生理过程的建模、检测和分类，还包括利用计算机系统的庞大基因数据库和蛋白质数据库实现基因组和蛋白质组的检索，同时分析微阵列芯片实验中细胞的信使 RNA（即 mRNA）的水

⊖ Susan Blanchard 和 Anne-Marie Stomp 也参与了部分撰写工作。——原文注

平。通过计算生物学方法的应用，可以仿真各种生理代谢途径，深入了解各种生理途径之间的相互作用，或者寻找治疗特定疾病的药物作用靶标。这些方法是连接纯理论知识与实验数据的纽带。计算生物学的最终目标就是获得细胞活动的各种信息，并通过这些信息的整合来揭示生物细胞的工作机制。

14.1.1 细胞生理过程的建模

细胞生理过程的建模可以分为两类，一是细胞内部调控机制的建模仿真，二是细胞内部新陈代谢途径的仿真建模。通常，细胞的这两种生理过程都与细胞的外部环境之间存在相互作用。例如，细胞的调控会受到来自邻近其他细胞、或者来自较远处组织、或者来自细胞所依附的胞外基质的刺激作用的影响。如代谢物的浓度会受到胞外环境中葡萄糖浓度的影响等。因此，细胞生理过程的这两类建模都需要考虑胞外因素。另外，建模时还需要考虑各种不同房室模型的特点（参见第 12 章），不同的房室模型可能对各种调控途径之间或者各种代谢物浓度之间的相互作用具有一定的约束。真核细胞建模时要建立细胞质、线粒体和细胞核等房室，可能还要考虑高尔基体和肌质网等房室。

14.1.2 细胞调控机制的仿真建模

细胞内部到处都存在调控。细胞膜上的受体与胞外物质相互作用后就会启动胞内的各种反应。有些受体的作用与 DNA 相关，会促进或抑制 mRNA 的产生，从而调控蛋白质的合成。这种受体可以位于细胞膜上或者细胞质中，通过一系列级联的蛋白质信使影响 mRNA 和蛋白质的生成，或者与第二信使产生作用，最终使细胞内的蛋白质浓度发生变化。其中某些蛋白质还可能诱发癌变，这类蛋白质受体被称为原癌基因（proto-oncogene），如酪氨酸激酶就是其中一例。另外，细胞核中的类固醇激素受体还可以穿过细胞核的质膜。有的受体可以促进各种代谢途径中底物的吸收，例如，葡萄糖转运蛋白 GLUT 受体可以促使细胞吸收血液中的葡萄糖。

14.1.3 细胞代谢途径的仿真建模

细胞的各种代谢途径使细胞能够利用葡萄糖、酮酸和某些氨基酸等各种分子中存储的能量，完成细胞分裂等不同的细胞生理功能。分子的能量存储于特定原子之间形成的化学键中。代谢途径包含所谓的“呼吸”过程，此过程利用氧分子的特性协助化学转变，从而释放“燃料”分子中的能量。不过，不通过氧气的作用，葡萄糖中的能量也可以释放，只是这种“无氧呼吸”的效率比较低。“无氧呼吸”释放能量的过程中会产生乳酸，乳酸会改变 pH 值，并对细胞内各种酶的活性产生不利的影响。

能量的释放和转化过程需要多种酶的参与，并有很多调控位点，相当复杂，整个过程还需要经过细胞内多个房室才能完成。例如，柠檬酸循环（即 Krebs 循环）一部分在细胞质中进行，另一部分则在线粒体中进行。因此，如果要仿真糖酵解途

径和 Krebs 循环，就必须建立细胞质和线粒体的模型，并且必须使 Krebs 循环产生的代谢产物能够进入线粒体，而线粒体中的三磷酸腺苷（即 ATP）则能够出来。

14.2　仿真建模步骤

无论是建立细胞的调控模型还是代谢模型，首先都必须明确建模的目标，只是为了认识细胞的生理过程呢？还是在新药的开发过程中需要识别关键的调控位点？或者是为进一步的实验工作提供信息和数据？这个问题很重要，因为它决定了如何建模以及如何设定模型的假设条件。这些假设条件可以影响模型仿真所需的软件计算平台的选择。如果用现成的软件包来实现仿真计算，那么，就要检查该软件包设计和开发时所做的设定是否与所需仿真的模型一致。

建模的目标还决定了模型仿真计算所需的运行时间，简单的模型在个人计算机或工作站上只需几分钟就可以完成，而较大规模的模型即使在巨型机或者 Linux 机群上也需要运行数小时，甚至数日。

14.2.1　仿真方法

计算模型可以用现成的高级软件包实现仿真，如：MATLAB、SIMULINK、V-Cell（网址为 http：//www.nrcam.uchc.edu/，见图 14.1）、E-Cell（网址为 http：//www.e-cell.org/）、Jig Cell（网址为 http：//jigcell.biol.vt.edu/）、LabVIEW 和 Microsoft Excel 等。也可以用 C/C++、Java 和 Basic 等程序设计语言编写仿真程

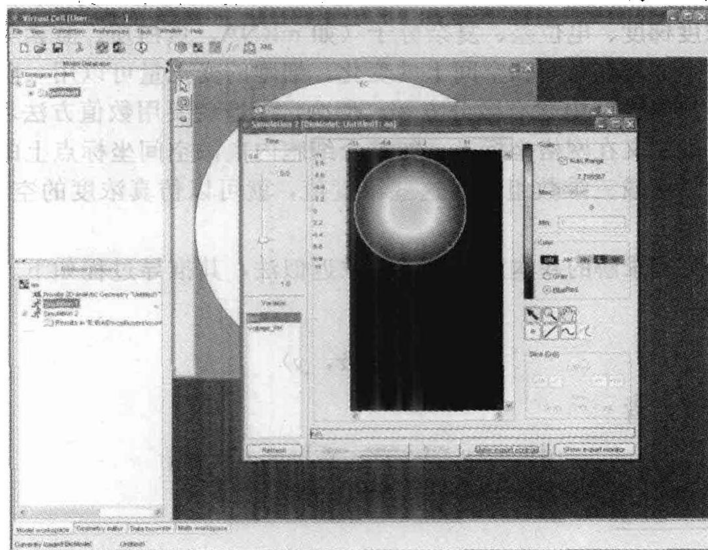


图 14.1　用于仿真细胞生理过程的软件包 V-Cell

注：图中的彩色圆圈显示了某特定时刻下物质向细胞内扩散的情况，时间可由用户设定。

序。高级软件包具备友好的界面环境，使用户可以快速创建模型并进行仿真计算，减少了调试时间，并可以利用软件现成的图形功能。有些软件包只需用户点击和拖拉各种图标，将它们放到工作空间中，就可以进行仿真。这类软件包的缺点是：有时不能仿真所需的模型特性，并且会增添很多额外开销，使复杂模型的仿真时间大大增加。如果使用程序设计语言自己编程仿真，其优点是：仿真速度快，灵活性大，可以根据需要仿真各种细节问题。但是，编写程序需要花费大量调试时间，并且还需要花力气开发仿真结果的输出显示；另外，还要求用户对程序设计语言有充分的了解和掌握，比用 V-Cell 或 E-Cell 等软件包的要求要高得多。不过，可以将高级软件包与用户自己用程序语言开发的模型结合起来；这样，既可以发挥编程语言的快速和灵活性，又可以利用软件包现成的图形显示功能。

计算模型由方程组构成，一般，各个模型方程描述了细胞或组织在生理调控过程中的各种变化量，如离子等代谢产物的浓度、胞内物质的流量等。这些变化量可以是空间变量，如细胞内各个房室之间的浓度梯度等；也可以是时间变量，如肌肉收缩时肌细胞内钙离子浓度的变化曲线等。这些变化通常用常微分方程或者偏微分方程来描述，不过，也可以用具有延时的布尔网络等其他方法来描述。模型的微分方程可以是存在解析解的可求解方程，也可以是只有数值解的复杂微分方程。

常微分方程的一个简单解析解就是指数函数，此函数的系数可以根据模型的特性来确定。

14.2.2 模型方程

正如 14.2.1 节所述，模型的方程主要包含了各种变化量，例如：酶的浓度、底物的浓度、浓度梯度、电位差、复杂分子（如 mRNA、蛋白质、激素等）的降解速率等，以及各个变量在时间和空间上的变化。细胞的变化量可以用一阶和二阶微分方程来描述。如果模型中含有微分方程，那么，一般就要用数值方法求其积分。模型的数值解一般定义在网格坐标上，如整个细胞内离散空间坐标点上的一组浓度值等。用一维、二维或三维数组储存这些浓度值，就可以仿真浓度的空间变化过程，并求取浓度梯度。

一阶微分方程求解的基本方法就是欧拉近似法，其推导过程如下。将微分方程写成如下形式：

$$\frac{\mathrm{d}y}{\mathrm{d}x}=f(x,\ y) \qquad (14.1)^{⊖}$$

根据导数的定义，有

$$\frac{\mathrm{d}y}{\mathrm{d}x}\approx\frac{\Delta y}{\Delta x}=\frac{(y_{n+1}-y_n)}{\Delta x}\approx f(x_n, y_n) \qquad (14.2)^{⊜}$$

⊖ 原文中此式为$\frac{\mathrm{d}f}{\mathrm{d}x}=f(x,\ y)$。——译者注

⊜ 原文中此式为$\frac{\mathrm{d}f(x)}{\mathrm{d}x}=\frac{\Delta f(x,y)}{\Delta x}=(f(x_2,y)-f(x_1,y))/\Delta x$。——译者注

改写此方程就可以得到欧拉法的计算公式：

$$y_{n+1}=y_n+\Delta x f(x_n,y_n) \tag{14.3}$$ ㊀

这种近似法的精度不高，对于某些方程还会导致不稳定解。由于欧拉法是利用函数在前一时刻的值（即 y_n）来估算当前时刻的值（即 y_{n+1}），因此，对于变化速度较快的函数，近似结果就不精确。如果用二阶或四阶龙格一库塔法（Runge-Kutta method）等其他算法计算导数，就可以得到比较稳定且精确的解。

二阶龙格一库塔法的近似公式是：

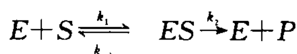
$$k_1=hf(x_n,y_n) \tag{14.4}$$

$$k_2=hf(x_n+\frac{1}{2}h,y_n+\frac{1}{2}k_1) \tag{14.5}$$

$$y_{n+1}=y_n+k_2+O(h^3) \tag{14.6}$$ ㊁

龙格一库塔法在步长的中点上估算函数值，因此，其计算结果可以较紧密地跟踪函数的变化过程。但是，如果函数在某个阶段的变化率有较大的变动，例如函数由缓慢变化突然变成快速变化，那么，上述这些近似算法都会出现问题。具有这种变化的微分方程被称为“刚性”方程。在仿真生物细胞信号传导途径时就会遇到刚性常微分方程。求解这种方程的一种常用方法是改变积分步长，当函数快速变化时，将积分步长 Δx 减小。很多高级仿真软件包允许用户根据仿真需要，选用不同的方法求解各种一阶微分方程。

Michaelis—Menten 方程（即米氏方程）是描述新陈代谢酶促反应动力学的基本一阶微分方程，它反映了产物生成速率与底物浓度之间的关系。在酶促反应中，酶 E 与底物 S 结合之后，就可以降低化学反应的活化能，从而加速底物向产物的转化。如果没有酶的作用，底物向产物的转化会非常慢，因此，通过降低活化能，酶增加了转化概率，提高了转化速度。每个酶促反应过程中，底物都有可能在没能转化成产物之前就与酶分离了，这种概率用常数 k_m 来表示。k_m 等于酶一底物复合物的分解速率与其生成速率之比，因此，它反映了酶一底物复合物的稳定性。Michaelis—Menten 方程假设相对于底物浓度而言，酶的浓度非常小。如果酶促反应为：



$$E+S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E+P$$

那么，其 Michaelis—Menten 方程为

$$\frac{\mathrm{d}[P]}{\mathrm{d}t}=V_{\max}[S]/(k_m+[S]) \tag{14.7}$$

其中

$$k_m=(k_{-1}+k_2)/k_1 \tag{14.8}$$

并且

$$V_{\max}=k_2[E_o] \tag{14.9}$$

㊀ 原文中此式为 $x_{n+1}=x_n+\Delta x f(x_n,y_n)$。——译者注

㊁ 原文中此式为 $x_{n+1}=x_n+k_2+O(h^3)$。——译者注

式中 V_{max}——最大转化速率，就是所有的酶都与底物结合的情况下的反应速率，它与酶一底物复合物的生成和分解速率无关；

$[E_o]$——酶的浓度；

$[S]$——底物的浓度；

k_m——产物的生成速度等于$\frac{1}{2}V_{max}$时的底物浓度；

$\frac{d[P]}{dt}$——产物增加的速率；

k_1——E 与 S 结合成 ES 的速率；

k_{-1}——ES 分解成 E 和 S 的速率；

k_2——ES 转化成 E 和产物 P 的速率。

Michaelis—Menten 方程可以反映底物的零阶和一阶两种动力学过程。当底物浓度较小时，随着底物浓度的增加，化学反应速度也增加，表现出一阶动力学过程。如果底物浓度最后增加到使系统中的酶都被饱和结合时，产物增加的速度就会减小，也就是$[P]$的二阶导数变为负数，产物的生成速率就趋近于 V_{max}。此时，方程表现出零阶动力学过程，产物的生成速率不再与底物的浓度成线性正比关系。这些情况成立的基本假设条件是化学反应的第二步是不可逆的，也就是系统中的酶不能将产物催化成底物。其实，在最终的不可逆反应之前可以存在多步化学反应过程，它们具有各自相应的反应速率。Michaelis—Menten 方程经过适当改进之后就可以描述这类反应。

在模拟生物量的空间变化时，要用到扩散方程，它可以描述一维、二维或三维空间中物质浓度随时间的变化。笛卡尔坐标系下的一维扩散方程是

$$\frac{\partial u}{\partial t}=D\frac{\partial^2 u}{\partial x^2} \tag{14.10}$$

三维扩散方程为

$$\frac{\partial u}{\partial t}=D\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}\right) \tag{14.11}$$

此方程可以用于仿真细胞膜上的跨膜浓度梯度，或者用于仿真在外周压力作用下生物组织内存在的扩散现象。

1. 细胞信号转导模型举例

如图 14.2 所示，细胞调控和代谢过程中普遍存在的一种机制就是磷酸化反应。在这种反应过程中，酶从一类称为激酶的蛋白质中接受磷酸基团（即 PO_4^{3-}）。一旦酶

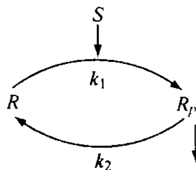


图 14.2 由刺激 S 诱发的酶的磷酸化反应示意图

注：磷酸化通过一种激酶完成，磷酸化酶 R_p 上的磷酸基团接下来可以转移给另一种酶。

被磷酸化之后，其行为就发生了变化，一般是被激活。这种磷酸化过程可以用微分方程来仿真，方程描述分子在非磷酸化状态与磷酸化状态之间的变化速率。注意，来回两个方向的反应速率常数是不同的，并且都与受到的刺激信号有关。

假设 R_p是磷酸化酶的含量（或者浓度），该磷酸化酶将参与其他反应过程。

描述 R_p变化的方程一般为

$$\frac{dR_p}{dt}=k_1S(R-R_p)-k_2R_p \tag{14.12}$$

式中 k_1——蛋白激酶将酶 R 磷酸化的速率；

k_2——磷酸酶将酶 R 去磷酸化的速率；

S——刺激信号；

R——非磷酸化形式的酶 R 的含量或浓度；

R_p——磷酸化形式的酶 R 的含量或浓度。

该方程描述了酶磷酸化和去磷酸化的速率，以及两种形式的酶的含量。方程的解是如图 14.3 所示的双曲线函数。

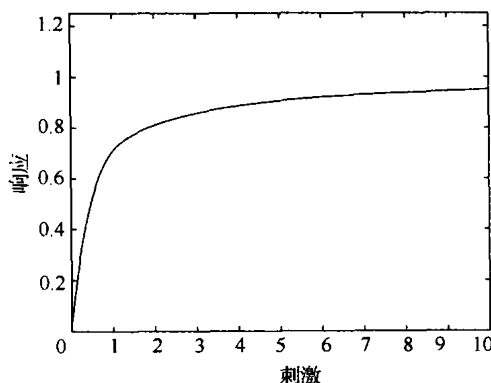


图 14.3 方程（14.12）定义的双曲线函数

磷酸化反应还有一种更精确的描述方程是

$$\frac{dR_p}{dt}=[V_1S(R-R_p)]/[k_1+R-R_p]-[V_2R_p]/[k_2+R_p] \tag{14.13}$$

式中 k_1——磷酸化酶 R_1的米氏常数；

k_2——磷酸化酶 R_2的米氏常数；

V_1——磷酸化的最大速率；

V_2——去磷酸化的最大速率；

S——刺激信号；

R——非磷酸化形式的酶含量或者浓度；

R_p——磷酸化形式的酶含量或者浓度。

该方程的 MATLAB 仿真程序如下：

```
function sigmoid
options = odeset('RelTol',le-4, 'AbsTol',[1e-4 le-4 le-5]);
Y = zeros(3,1); % a column vector
[t,Y] = ode23@s_model,[0 100],[0 0 0],options);
figure, semilogx(t,Y(:,1),'k-'); % Plots fig. 14.4
axis([0 100 0 1.25]);
xlabel('Stimulus')
ylabel('Response')
figure, semilogx(t,Y(:,1),'k-',t,Y(:,2), 'k--'); % Plots fig. 14.6
axis([0 100 0 1.25])
```

```
xlabel('Stimulus')
ylabel('Response')
figure, semilogx(t,Y(:,1),'k-',t,Y(:,2), 'k--',t,Y(:,3),'k.') % Plots fig. 14.8
axis([1 100 0 1.25])
xlabel('Stimulus')
ylabel('Response')

function dy = s_model(t,y)
dy = zeros(3,1);
dy(1) = (0.05*t*(1 - y(1)))/(0.2 + (1 - y(1))) - (0.50*y(1))/(0.10 + y(1));
dy(2) = (0.87*y(1)*(1 - y(2)))/(0.10 + (1 - y(2))) - (0.10*y(2))/(0.1 + y(2));
dy(3) = (0.85*y(2)*(1 - y(3)))/(0.08 + (1 - y(3))) - (0.08*y(3))/(0.1 + y(3));
```

程序运行的输出结果如图 14.4 所示。由于这些酶的反应遵循 Michaelis--Menten 动力学过程，因此，这个方程更切合实际情况。

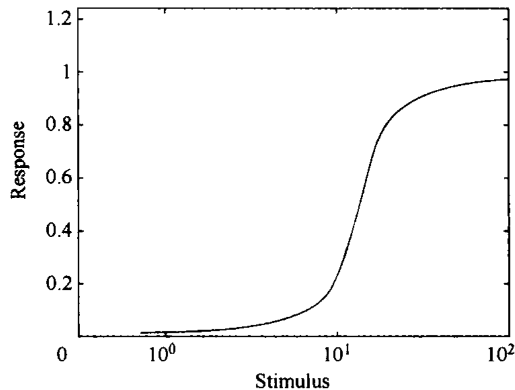


图 14.4 由方程（14.13）定义的 S 型曲线

如果一种酶的磷酸化会导致另一种酶的磷酸化，那么信号就会沿着反应链传递下去。图 14.5 所示的反应链中就含有两个磷酸化反应，可用以下两个方程描述：

$$\frac{dR_p}{dt}=[V_1S(R-R_p)]/[k_1+R-R_p]-[V_2R_p]/[k_2+R_p] \tag{14.14}$$

$$\frac{dR_{1p}}{dt}=[V_3R_p(R_1-R_{1p})]/[k_3+R_1-R_{1p}]-[V_4R_{1p}]/[k_4+R_{1p}] \tag{14.15}$$

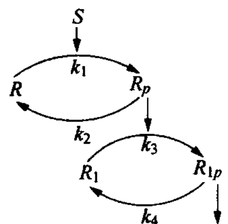


图 14.5 含有两个磷酸化反应的激酶级联反应链

该方程组的 MATLAB 仿真程序如下：

```
function sigmoid
options = odeset('RelTol',le-4, 'AbsTol',[1e-4 le-4 le-5]);
Y = zeros(3,1); % a column vector
[t,Y] = ode23(@s_model,[0 100],[0 0 0],options);
figure, semilogx(t,Y(:,1),'k-'); % Plots fig. 14.4
axis([0 100 0 1.25]);
xlabel('Stimulus')
ylabel('Response')
```

```
figure, semilogx(t,Y(:,1),'k-',t,Y(:,2), 'k–'); % Plots fig. 14.6
axis([0 100 0 1.25])
xlabel('Stimulus')
ylabel('Response')
figure, semilogx(t,Y(:,1),'k-',t,Y(:,2), 'k–',t,Y(:,3),'k.') % Plots fig. 14.8
axis([1 100 0 1.25])
xlabel('Stimulus')
ylabel('Response')

function dy = s_model(t,y)
dy = zeros(3,1);
dy(1) = (0.05*t*(1 - y(1)))/(0.2 + (1 - y(1))) - (0.50*y(1))/(0.10 + y(1));
dy(2) = (0.87*y(1)*(1 - y(2)))/(0.10 + (1 - y(2))) - (0.10*y(2))/(0.1 + y(2));
dy(3) = (0.85*y(2)*(1 - y(3)))/(0.08 + (1 - y(3))) - (0.08*y(3))/(0.1 + y(3));
```

程序运行的输出结果如图 14. 6 所示。

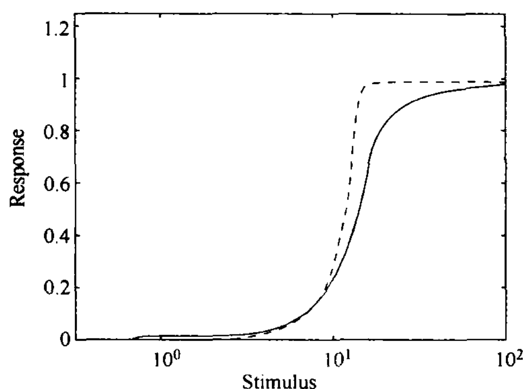


图 14. 6　方程（14. 14）和（14. 15）的解

注：图中所示的两条 S 型曲线中，实线表示刺激诱发的第一个磷酸化反应，虚线表示由第一个磷酸化反应诱导的第二个磷酸化反应。

图 14. 7 所示的反应链中含有 3 个磷酸化反应，因此，以上仿真模型中要再添加第三个磷酸化反应的方程，即：

$$\frac{dR_{2p}}{dt}=[V_5R_p(R_2-R_{2p})]/[k_5+R_2-R_{2p}]-[V_6R_{2p}]/[k_6+R_{6p}] \quad (14.16)$$

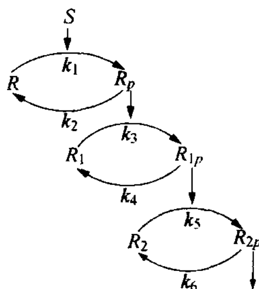


图 14. 7　含有 3 个磷酸化反应的激酶级联反应链

其 MATLAB 仿真程序如下：

```
function sigmoid
options = odeset('RelTol',le-4, 'AbsTol',[1e-4 le-4 le-5]);
Y = zeros(3,1); % a column vector
[t,Y] = ode23@s_model,[0 100],[0 0 0],options);
figure, semilogx(t,Y(:,1),'k-'); % Plots fig. 14.4
axis([0 100 0 1.25]);
xlabel('Stimulus')
ylabel('Response')
figure, semilogx(t,Y(:,1),'k-',t,Y(:,2), 'k--'); % Plots fig. 14.6
axis([0 100 0 1.25])
xlabel('Stimulus')
ylabel('Response')
figure, semilogx(t,Y(:,1),'k-',t,Y(:,2), 'k--',t,Y(:,3),'k.') % Plots fig. 14.8
axis([1 100 0 1.25])
xlabel('Stimulus')
ylabel('Response')

function dy = s_model(t,y)
dy = zeros(3,1);
dy(1) = (0.05*t*(1 - y(1)))/(0.2 + (1 - y(1))) - (0.50*y(1))/(0.10 + y(1));
dy(2) = (0.87*y(1)*(1 - y(2)))/(0.10 + (1 - y(2))) - (0.10*y(2))/(0.1 + y(2));
dy(3) = (0.85*y(2)*(1 - y(3)))/(0.08 + (1 - y(3))) - (0.08*y(3))/(0.1 + y(3));
```

程序的运行结果如图 14.8 所示，磷酸化反应曲线的特性由各个速率常数决定。

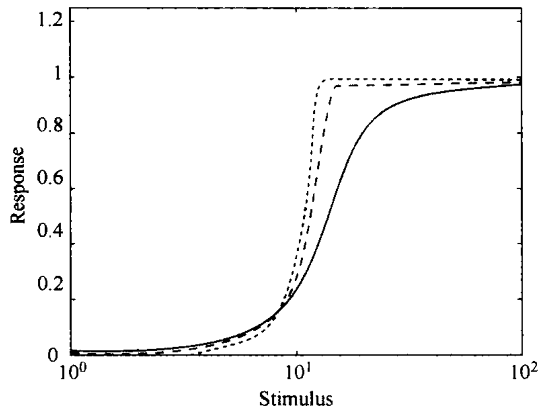


图 14.8 方程（14.14）～（14.16）的解

注：图中所示的 3 条 S 形曲线中，实线表示刺激诱发的第一个磷酸化反应；虚线表示由第一个磷酸化反应诱导的第二个磷酸化反应；点线表示受到第二个磷酸化反应控制的第三个磷酸化反应。程序中所采用的磷酸化反应的常数并不是实际 MAPK 反应链的数据，只是用于示意。

这个含有 3 个磷酸化反应的级联反应其实是丝裂原激活蛋白激酶（mitogen-activated protein kinase，MAPK）信号通路的简化模型（见图 14.9）。MAPK 参与许多细胞调控过程，包括细胞的生长、分化、炎性反应和凋亡等。MAPK 通路是一个由蛋白激酶组成的具有高度保守性的反应链，它接收来自各种整合素、G 蛋白耦联型受体（GPCRs）和酪氨酸激酶受体（RTKs）等各种刺激源的信号。其中，通过整合

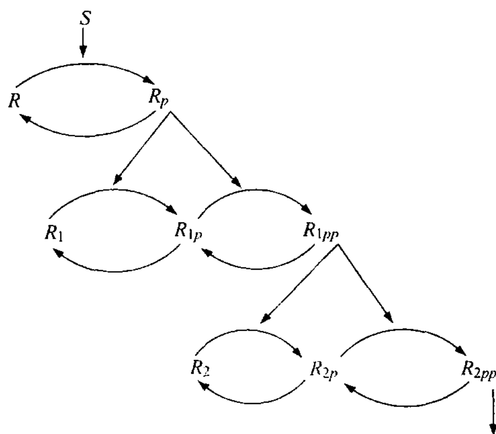


图 14.9　MAPK 信号转导通路

注：该信号通路在很多物种中都具有高度的保守性，它保证了某些特定细胞生理过程的严格控制。这种信号转导级联反应包含了多个磷酸化过程，通过酶来传递信号。发生第一个磷酸化反应的是响应刺激信号的 MAPKK 激酶 R；该激酶磷酸化之后将 MAPK 激酶 R_1 磷酸化两次；接着，再将 R_2 磷酸化。

素，MAPK 级联通路可以接受胞外基质作用于细胞膜上的应力和应变等信号；G 蛋白耦合受体则通常会传递来自胞外非固醇类激素的有关远处细胞组织状态的信号；而酪氨酸激酶受体则提供来自胞外生长因子（如表皮生长因子 EGF 等）的信号。MAPK 级联反应将这些信号与其他信号一起进行整合，用于决定是否启动基因表达，并由此来触发和控制细胞的调控过程。细胞中已经发现多种不同的 MAPK 信号转导通路，如 ERK1/2、ERK5、p38 和 JNK 等。Bhalla 和 Iyengar（1999）、Huang 和 Ferrell（1996）、Kholodenko（2002），以及 Shvartsman（2001）等人的论文提出了一种仿真 MAPK 通路的简化模型。在这种模型中，刺激信号引发最初的磷酸化反应，然后信号沿着反应链传递下去。在传递过程中，刺激的响应表现为开关式的双稳态反应。

2. 蛋白质合成模型举例

本例利用磷酸化方程仿真信使 RNA（即 mRNA）的转录以及转录所产生的 mRNA 再经过翻译合成蛋白质的过程。模型方程如下，其中，第一个方程描述启动子的反应，启动子被磷酸化之后，转录就开始了。第二个方程描述转录开始后 mRNA 不断增加并随即开始衰减的过程。第三个方程描述 mRNA 翻译并合成蛋白质的过程，蛋白质浓度先增加而后即开始衰减。

$$\frac{\mathrm{d}R_p}{\mathrm{d}t}=[V_1R_p(R-R_p)]/[k_1+R-R_p]-[V_2R_p]/[k_2+R_p] \tag{14.17}$$

$$\frac{\mathrm{d}[\mathrm{mRNA}]}{\mathrm{d}x}=k_3*[R_p]-k_4*[\mathrm{mRNA}] \tag{14.18}$$

$$\frac{\mathrm{d}[\text{蛋白质}]}{\mathrm{d}x}=k_5*[\mathrm{mRNA}]-k_6*[\text{蛋白质}] \tag{14.19}$$

式中　k_3——mRNA 生成的速率；

k_4——mRNA 降解的速率；

k_5——蛋白质合成的速率；

k_6——蛋白质降解的速率。

其 MATLAB 仿真程序如下：

```
function simulation
options = odeset('RelTol',le-4, 'AbsTol',[1e-4 le-4 le-5]);
Y = zeros(3,1); % a column vector
[t,Y] = ode23(@simmodel,[0 100],[0 0 0],options);
figure, plot(t, Y(:,1),'k-',t,Y(:,2), 'k--',t,Y(:,3),'k.');
xlabel('time')
ylabel('Concentration')

function dy = simmodel(t,y)
dy = zeros(3,1);% a column vector

k11 = 0.15; % Scaling purposes only
k1 = 0.2;
k2 = 0.1;
k3 = 0.1;
k4 = 0.1;
k5 = 0.1;
k6 = 0.01;
v1 = 0.6;
v2 = 0.15;

if t <2 % Used to model a limited stimulus
s = 5; % Length of stimuli
else( t>2)
s = 0;
end
dy(1) = (v1*k11*s*(1 - y(1)))/(k1 + (1 - y(1))) - (v2*y(1))/(k2 + y(1));
dy(2) = k3*y(1) - k4*y(2);
dy(3) = k5*y(2) - k6*y(3);
```

程序的运行结果如图 14.10 所示，mRNA 衰减的速率比蛋白质衰减的速率要快。

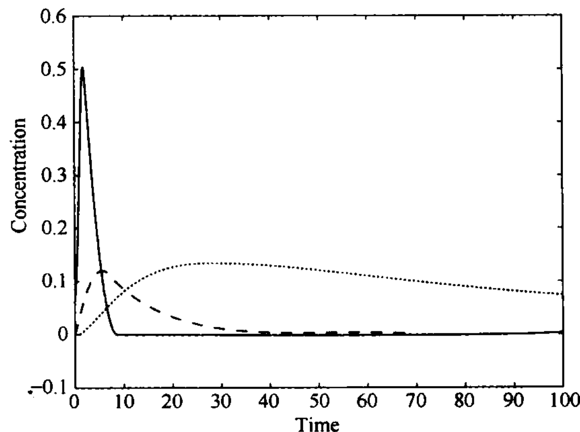


图 14.10 由启动子引导转录并继而引发蛋白质翻译过程的简化模型

注：实线表示磷酸化启动子的浓度，虚线表示 mRNA 的浓度，点线是蛋白质产物的浓度。

该模型的仿真结果显示了蛋白质产物的初始合成过程。其实，在细胞生理过程中蛋白质的合成需求是周期性变化的。如果要使仿真得到的蛋白质浓度保持在细胞生存所需的浓度上，那么，就必须根据蛋白质降解速率，周期性地施加刺激信号，诱导蛋白质的周期性合成过程。图 14.11 显示了在周期性刺激信号作用下蛋白质浓度的仿真结果。

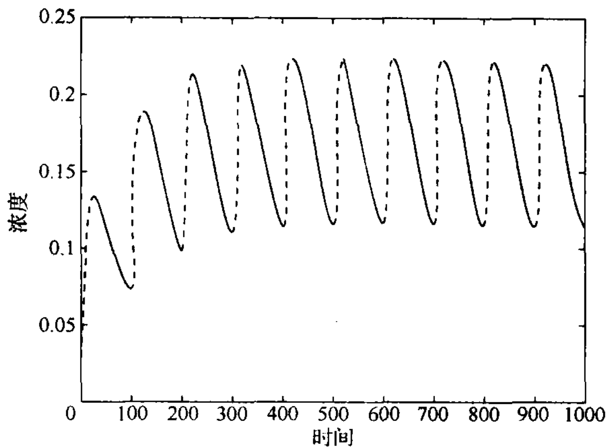


图 14.11 蛋白质浓度变化

注：简单修改 MATLAB 程序，每 100 个时间单位施加一个周期性刺激信号，这样，蛋白质的合成就可以补偿蛋白质的降解，使蛋白质浓度维持在所需的最低浓度之上。

14.3 生物网络

上述两个例子其实只显示了代谢途径一连串反应中的少数几步反应。生物网络是细胞中完成某种特定生物功能的复杂途径。这种途径包括很多反应步骤，需要多种不同蛋白质的参与，并且可能在很多个点上与其他途径有连接。通过这些连接点，某个途径可以接收其他途径传导过来的生物分子，也可以向其他途径输出生物分子。糖酵解、柠檬酸循环和氧化磷酸化等都是不同的生物网络。这种生物网络分析方法面临的难题是各个网络之间都存在相互作用，它们通过底物、产物以及维持细胞状态的 ATP、NADH、NAD 等各种分子联系在一起。这些相互联系大大增加了生物网络分析的复杂性。

如图 14.12 所示的生物网络，其中的每一步反应都可以用单个化学方反应式来描述，表示底物与产物之间的关系，每种产物随后又成为代谢途径中下一个化学反应的底物。于是，就可以写出一组线性方程来描述整个系统的化学计算，然后将方程组合成计算矩阵，通过求解矩阵就可以分析系统的各种特性。如基因网络

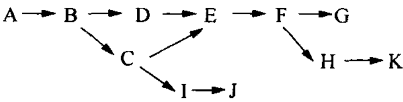


图 14.12 简单生物网络的示意图

就表示了细胞生理过程的一种调控机制。通过仿真多个相互作用的代谢途径，就可以研究特定蛋白质的失活或者活性降低所产生的影响。如果生物网络反映的是疾病引起的某些变化，那么，部分代谢途径的失活所产生的结果就可以反映特定蛋白质的竞争性抑制药物的作用效果。在京都基因与基因组百科全书网站（Kyoto Encyclopedia of Genes and Genomes）（http：//www. genome. jp/kegg/）上可以找到大量代谢途径和调控途径的数据库。由于存在冗余途径、不同途径之间的相互作用以及细胞对于各种途径的广泛调控，无论是代谢途径还是控制途径，生物网络都具有相当高的复杂性。

14.4 复杂性理论概述

复杂性理论研究各种系统之间的相互作用关系，它用于生物系统研究时，可以认识和表征维持生命系统的各种不同生物功能的机制。复杂性研究基于系统的两方面行为，一是系统对于特定刺激信号的响应，二是各个子系统的作用以及子系统之间相互作用所产生的结果。通过对于复杂系统的研究，既可以认识各种系统的共性，也可以确定维持系统生存所需的条件。

一般，复杂系统具有以下特性：

1）大量局部子系统之间存在着丰富的相互作用；

2）子系统之间的相互作用可以是抑制性的、刺激性的、竞争性的、或者协同性的，并且还可能与环境条件或者其他子系统有关；

3）由于存在子系统之间的相互作用，就可能表现出各种系统行为的变化，即自由度很高；

4）如图 14.13 所示，系统可能呈现的各种行为状态会随时间和空间变化，即具有时空自由度；

5）系统的整体行为大部分由各个子系统之间的相互作用决定，受各个子系统自身行为状态的影响则比较少；

6）在各个子系统之间的相互作用以及子系统自身特性这两方面因素的影响下，整个系统的行为表现出非线性；

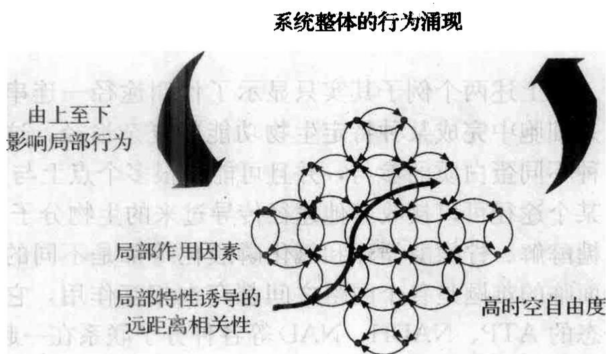


图 14.13 各种尺度之间相互作用示意图以及系统行为的涌现

7）各个子系统并不是受某个中心过程的调控，而是各个子系统之间的相互作用上都广泛受到调控（即自组织）；

8）系统整体行为的涌现取决于自组织；

9）复杂系统能够适应较小的环境变化并保持稳定；

10）复杂系统具有临界现象，也就是，它可以产生一种状态到另一种状态的突变。

如图 14.14 所示，复杂系统由大量子系统组成，这些子系统之间存在很多相互作用途径，相互影响，并可能影响其他子系统的输出。复杂系统具有很多维自由度，存在空间分布，这使得系统的作用具有时间和空间自由度。这表明，由刺激“诱发”产生的各种事件会随时间和空间变化。因此，系统所呈现的整体行为较多取决于各个子系统之间的相互作用，而不是子系统各自的行为；也就是，系统整体行为是从各种相互作用中“涌现”出来的，它实际上是由时间和空间上分布的相互作用建立起来的，从而导致了分布式的“自组织”。系统的稳定性和适应性与分布式自组织相关。子系统之间的相互作用和子系统各自的行为都是非线性的。

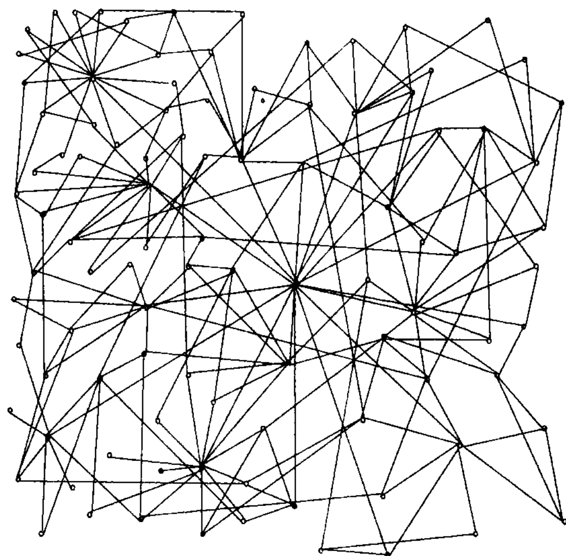


图 14.14　复杂系统相互作用图的一个例子

注：图中每个节点可以代表蛋白质、转录位点等。每个节点都与其他节点之间具有相互作用，但没有节点与所有其他节点都有相互作用。分布于各个节点之间的相互作用决定了激活哪些过程，并确定了系统的行为。

14.4.1　代谢网络和调控网络的复杂性

虽然利用线性微分方程组可以独立建模仿真细胞的各种代谢途径和调控途径，获得有关生物功能的知识；但是，这种简化方法并不能全面揭示机体的复杂性。细胞的各个代谢途径和调控途径都与其他途径之间存在相互作用，这些相互作用可以导致反应物进入其他途径并产生其他作用。各种调控途径之间的联系可以将细胞状态转导给其他调控过程，在某些情况下，代谢途径也具有这种功能。这种细胞通信可以传递有关细胞周围环境的信息（如离子浓度和旁分泌或自分泌反应等）、有关远

处组织的状态信息（如内分泌）、或者有关细胞自身的状态信息。由于细胞内外的状态信息都传递给了其他途径，这种机制是一种分布式“决策”方法，它决定哪些途径为主动或者被动。由主动和被动途径形成的细胞行为会引导细胞在机体中完成特定的任务，这种系统行为被定义为适应性，它是复杂性的重要特性之一。分布式相互作用是行为涌现的一个机制，也是复杂性理论的特征之一。行为涌现这个性质的通俗说法就是“整体大于部分之和”，专业的说法则是“多个途径和子系统之间的相互作用产生了系统行为，这种行为并不能通过各个途径和子系统的分立研究获得”。

细胞内 ATP 的生产就是多途径通信的一个例子。ATP 是细胞的能量货币，细胞内 ATP 的浓度决定了细胞所处的能量状态。高浓度 ATP 表示细胞处于高能量状态，它具有较高的概率提供磷酸基团（即 PO_4^{3-}），用于激活某些蛋白，然后使某些特定的代谢途径或控制途径激活或者失活。类似的例子还有 GTP 和 cAMP 等细胞内信使。

不过，在这种多途径复杂关系形成的分布式“决策”机制作用下，细胞也必须具备稳定性。如果环境的小扰动使细胞状态发生了小变化，稳定性就会起作用，如细胞外钠或钾浓度产生小变动时细胞仍然能够存活。但是，较大的环境变动就有可能导致系统的灾难性变化，如细胞可能死亡。对于非稳定系统，即使很小的扰动都会使系统产生不可预测的灾难性变化。

稳定性对于细胞的新陈代谢、调控和繁殖等功能尤其重要。不过，虽然稳定性对于维持特定机体的生存很重要，但是，进化却需要变动。因此，虽然细胞的代谢和分布式调控必须稳定，但在细胞周期的 S 期的 DNA 复制过程中存在非零失误率却是有利的。这样，DNA 或 RNA 的变异就可以引起蛋白质等功能元素的变化，从而提高细胞的适应能力。于是，机体就可以与环境抗争，最终适应环境，并根据它获得的适应性繁衍。镰刀状细胞贫血症可能就是通过这种变化过程形成的。

在系统建模过程中，通常都假设生理过程是线性的，这是为了简化求解，所得到的解只是真实系统的近似解。大多数系统实际上都是非线性的，这是复杂性的又一个特性。

尽管本文所述只是细胞行为层次上的复杂性，但是，实际上，从亚细胞到群体和生态系统，复杂性存在于生命系统的所有层次。细胞系统通过各种不同代谢途径和调控途径之间的相互作用呈现出复杂性，这些途径使细胞能够对内部或外部刺激作出响应并保持活力，使细胞具有适应性和稳定性。但是，细胞系统所表现的行为与独立分析各个子系统所得到的预测结果是不同的，它呈现的是行为的涌现。

总之，计算机可以用于细胞内特定蛋白质功能的识别、mRNA 表达水平的分析和聚类以及代谢和调控生物网络的建模，这 3 个方向的研究工作正在不断揭示细胞功能的新机制，识别出疾病治疗的药物靶标。这些研究方法结合起来分析现有的实验数据，可以建立有关细胞信息转导途径的理论，然后再用实验数据验证理论假设。通过各种代谢和调控途径的仿真，计算机已经可以模拟细胞的某些复杂性。随着计算机硬件功能的增强、各种仿真算法的开发，以及分子生物学提供的有关细胞功能

相互作用的知识和定量数据的日益增加，计算机模型的仿真将继续不断揭示生物机体更大的复杂性。

习　题

1. Michaelis—Menten 方程所描述的酶促反应动力学可以用一组一阶微分方程表示，其中各个微分方程表示每种物质的变化率等于速率常数与每种反应物浓度的乘积。请写出描述酶促反应的 4 个微分方程，并说明方程中每一项的含义。

2. 如果酶促反应中 ES 到 $E+P$ 的转变是可逆的，请写出 Michaelis—Menten 方程的 4 个一阶常微分方程。

3. 磷酸化信号转导模型的 S 形曲线的特征与函数的常数和刺激信号有关。请改变常数，观察曲线的变化情况。当常数的取值在什么范围时，可以使曲线变成双曲线的？

4. MAPK 级联反应是非常重要的细胞信号转导途径，该反应链的变化会导致哪些疾病的产生？

5. 本章所示的 MAPK 级联模型没有考虑反馈作用。哪些细胞机制可能对此反应链产生负反馈或者正反馈调控？

6. 炎症会激活 NF-κB 信号传导通路，该途径已有几种仿真模型（Hoffmann 等，2002；Lipniacki 等，2004）。该通路中激活了哪些重要的蛋白质？它是如何受到调控的？其中存在反馈作用吗？激活 NF-κB 通路的结果是什么？哪些疾病是由 NF-κB 通路的变化引起的？

7. 细胞分裂模型显示了细胞内蛋白质浓度的振荡效应。该模型中包括哪些蛋白质？磷酸化在此起了什么作用？各种蛋白质是如何调控细胞周期的？

8. 刺激频率、合成速度和降解速率，这些因素一起决定了细胞内蛋白质的浓度。请修改本章的蛋白质合成仿真程序，增加仿真时间，即：

（1）将 Simulation 函数中的相应语句改为：

```
[t, Y] =ode23 (@simmodel, [0 1000], [0 0 0], options);
```

（2）将 Simmodel 函数中的相应语句改为：

```
if mod (t, 100) <2          %Used to model a limited stimulus
s=5;
else (mod (t, 100) >2)
s=0;
end
```

并且，增加或减小刺激频率、合成速率和降解速率，考察仿真结果的变化，并与图 14.11 进行比较。

9. 针对磷酸化级联反应的例子，说明什么是 Goldbeter-Koshland 函数和超敏感性？它们如何影响磷酸化反应链？

10. 本章磷酸化级联反应的例子中假设细胞处于高能量状态。细胞的能量状态是由什么决定的？如果细胞处于低能量状态，各个反应曲线会如何变化？

11. 磷酸化级联反应也可以用一组一阶常微分方程来描述，只是其中某些方程用于表示输入刺激信号。请写出反应链中每种反应物分子的常微分方程。注意，每种磷酸化的分子与非磷酸化分子是不同的。

12. 细胞的正常生理过程和病理过程都存在涌现行为，请说明涌现行为是如何导致疾病的产生的。

13. 磷酸果糖激酶 PFK-1 是细胞的一种代谢激酶，它在哪些代谢途径中起作用？（请参考东京基因与基因组百科全书）。磷酸化可以调控 PFK-1，那么，磷酸化的 PFK-1 怎样使某些代谢途径激活和失活？它与细胞的能量状态又有什么关系？还有其他哪些分子可以调控 PFK-1 的活性，是如何调控的？

14. 请举例说明具有如下控制功能的细胞途径：（参见 Tyson 等，2003）：

a. 前馈回路；

b. 完全的适应性；

c. 正反馈；

d. 负反馈；

e. 底物消耗振荡器；

f. 激活剂-抑制剂振荡器。

15. 镰刀状细胞贫血症被认为是个体变异引起的适应性增强的结果，即机体对于特定环境条件的抵抗。导致这种选择的环境因素是什么？其他还有什么疾病也会通过增强个体变异的适应性来抵抗这种环境条件？

推荐阅读资料

Allman, E.S. and Rhodes, J.A. (2004). *Mathematical Models in Biology.* Cambridge, Cambridge, U.K.

Bhalla, U.S. and Iyengar, R. (1999). Emergent properties of networks of biological signaling pathways. *Science* **283**, 381–387.

Fall, C.P., Marland, E.S., Wagner, J.M. and Tyson, J.J. (Eds.). (2002). *Computational Cell Biology.* Springer-Verlag, New York.

Hoffmann, A., Levchenko, A., Scott, M.L. and Baltimore, D. (2002). The IκB-NF-κB signaling module: Temporal control and selective gene activation. *Science* **298**, 1241–1245.

Huang, C.Y. and Ferrell, J.E. (1996). Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc. Nat. Acad. Sci. USA* 93, 10078–10083.

Keener, J. and Sneyd, J. (1998). *Mathematical Physiology.* Springer-Verlag, New York.

Kholodenko, B.N. (2000). Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur. J. Biochem.* **267**, 1583–1588.

Nelson, D.L. and Cox, M.M. (2005). *Lehninger Principles of Biochemistry.* W.H. Freeman, New York.

Lipniacki, T., Paszek, P., Brasier, A.R., Luxon, B. and Kimmel, M. (2004). Mathematical model of NF-κB module. *J. Theoret. Biol.* **228**, 195–215.

Lodish, H. et al. (1999). *Molecular Cell Biology, W.H. Freeman & Co.*, New York.
Markevich, N.I., Hoek, J.B. and Kholodenko, B.N. (2004). Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *J Cell Biol.* **164**, 353–359
Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (2002). *Numerical Recipies in C++: The Art of Scientific Computing*, 2nd Ed. Cambridge Univ. Press, Cambridge, U.K.
Sauro, H.M. and Kholodenko, B.N. (2004). Quantitative analysis of signaling networks. *Prog. Biophys. & Molecular Biol.* **86**, 5–43.
Shvartsman, S.Y. et al. (2001). Autocrine loops with positive feedback enable context-dependent cell signaling. *Am. J. Physiol. Cell Physiol.* **282**, C545–C559.
Stephanopoulos, G.N., Aristidou, A.A. and Nielsen, J. (1998). *Metabolic Engineering Principles and Methodologies,* Academic, San Diego, CA.
Tyson, J.J. (1991). Modeling the cell division cycle: cdc2 and cyclin interactions. *Proc. Nat. Acad. Sci. USA* **88**, 7328–7332.
Tyson, J.J., Chen, K.C. and Novak, B. (2003). Sniffers, buzzers, toggles, and blinkers: Dynamics of regulatory and signaling pathways in the cell. *Curr. Op. Cell Biol.* **15**, 221–231.
Weng, G., Bhalla, U.S. and Iyengar, R. (1999). Complexity in biological signaling systems. *Science* **284**, 92–96.
Wolf, D.M. and Arkin, A.P. (2003). Motifs, modules and games in bacteria. *Curr. Op. Microbiol.* **6**, 125–134.

第 15 章　放射成像

Joseph Bronzino 博士，专业工程师

本章目录

本章主要内容：

- 放射性活度的基本概念；
- 医学成像的两种电离辐射检测方法：一种是将放射性药物引入人体，使人体变成辐射源，然后，检测其所发出的辐射并用于成像；另一种是使用体外辐射源发出的射线，让其穿过人体，检测透射射线并用于成像。

15.1　绪论

纵观整个医学的发展历史，各种疾病诊断技术的开发一直是医学界的主要奋斗目标之一，人们不断地努力寻找能够在尽可能小的创伤状况下“看见”人体内部组织结构的各种方法。然而，直到20世纪后半期，医学史上才出现了能够提供病人体

内正常组织和病变组织影像的仪器设备。如今的现代化成像设备基于多种不同的物理原理，如 X 射线学、核物理学、光学、超声学等，并结合最先进的计算机技术和数据处理技术，它们不仅在病情监测中发挥了极其重要的作用，而且正在不断推进医疗保健系统的变革。

一个多世纪以前的 1895 年，首届诺贝尔奖获得者，物理学家伦琴（Wilhelm Conrad Roentgen），发现了一种新的放射线，即 X 射线。从此诞生了一门新的医学学科，即放射医学，出现了医学成像产业。刚开始时，这种放射成像系统的功能并不强，只能显示对比度较大的骨骼以及泌尿系统和消化系统等人体结构。20 世纪 70 年代以后，医学成像技术的迅速发展，尤其是计算机的应用，引起了 X 射线医学成像技术的变革。继伦琴之后，AllanMacleod Cormack 和 Godfrey Newbold Hounsfield 取得了同样令人振奋的成就。当时，Cormack 在 Tufts 大学物理系工作，而 Hounsfield 则在英国 EMI 有限公司研究所工作，两人不约而同地产生了同样的研究思路，但完成的工作却截然不同。Cormack 在他那篇 1963 年发表的著名论文中优雅地证明了图像重建的数学理论。不到 10 年，Hounsfield 就令人难以置信地展示了世上第一台具有高对比度和分辨率的医用数字化轴向断层成像设备。为了表彰他们两人在开发计算机断层成像技术中的突出贡献，Cormack 和 Hounsfield 共同获得了 1979 年诺贝尔物理学与医学奖。他们的发明所带来的活力至今长盛不衰。

能够产生离子的辐射被称为电离辐射。根据辐射是由体内还是体外产生的，医学成像检测的电离辐射可以分为两类：一类是将放射性药物引入人体，使人体成为辐射“源”之后所产生的电离辐射；另一类是体外产生的电离辐射，穿过人体后，被放在人体另一侧的检测仪采集的透射辐射。

相应地，医用放射成像设备也有两类。一类设备常用于细胞和人体器官的功能研究，它们通过检测被测对象体内放射物所发射的 γ 射线或者 X 射线，采集与生理功能或者解剖结构相关的辐射数据，用于建立各种图像。90%以上的核医学诊断系统都使用这种激发与检测相结合的放射性核素成像技术。另一类是 X 射线机及其计算机断层成像系统，它们检测的则是穿过人体的外源性射线。这两类设备及其使用都属于医院的放射科。本章将介绍这两类放射成像设备的基本工作原理。

15.2　放射性核素成像技术

核医学作为医学的一个重要分支，其主要目标之一就是利用放射成像技术辅助医生作出正确的诊断。核医学是第二次世界大战期间原子时代核物理技术发展的产物，如今已经成为各种特殊生理疾病检测和治疗的有效方法。

核医学领域的发展是物理学新理论在医学中应用的一个典范，它体现了临床医生与物理学家之间的又一次成功联合。该领域已经发展成为一个拥有自己的理论和方法的交叉学科，其主要研究方向包括：

1）开发人体可以安全摄入的放射性示踪剂（即放射性药物），研究正确使用这

些放射性药物的方法；

2）设计制造核医疗仪器设备，用于检测并显示放射性核素的活度；

3）研究放射性示踪剂的活度与特定生理过程之间的相关性。

为了使读者能够更好地了解放射性核素成像的基本原理，下面首先介绍物质的放射性及其检测方法以及用于监测放射性物质活度的仪器设备。

15.2.1 基本概念

1895年，伦琴发现了一种极具穿透力的新射线，即X射线，并由此开辟了一个崭新的科学研究领域。X射线与无线电波和可见光光波一样，是一种电磁辐射，只是它们的频率和波长有所不同。表15.1列出了各种电磁辐射的波谱。可见，X射线的波长范围一般为0.01～100 nm，比无线电波和可见光的波长都要短很多。

在伦琴发现X射线事件的激励下，法国物理学家贝克勒尔（Henri Becquerel）经过深入研究发现，人们熟悉的荧光或磷光物质所产生的辐射其实与X射线相似。1896年，贝克勒尔宣布，无论是否属于荧光物质，某类铀盐也具有放射性，它能够发射具有穿透力的射线。这些令人吃惊的研究结果向世界展示了一种出人意料的物质新特性。

表15.1 电磁辐射波谱

能量/eV	频率/Hz		波长/m
4×10^{-11}	10^{4}		10^{4}
4×10^{-10}	10^{5}	AM无线电波	10^{3}
4×10^{-9}	10^{6}		10^{2}
4×10^{-8}	10^{7}	无线电短波	10^{1}
4×10^{-7}	10^{8}	FM无线电波和电视信号	10^{0}
4×10^{-6}	10^{9}		10^{-1}
4×10^{-5}	10^{10}	微波和雷达	10^{-2}
4×10^{-4}	10^{11}		10^{-3}
4×10^{-3}	10^{12}	红外线	10^{-4}
4×10^{-2}	10^{13}		10^{-5}
4×10^{-1}	10^{14}	可见光	10^{-6}
4×10^{0}	10^{15}	紫外线	10^{-7}
4×10^{1}	10^{16}		10^{-8}
4×10^{2}	10^{17}		10^{-9}
4×10^{3}	10^{18}	X射线	10^{-10}
4×10^{4}	10^{19}		10^{-11}
4×10^{5}	10^{20}		10^{-12}
4×10^{6}	10^{21}	γ射线	10^{-13}
4×10^{7}	10^{22}	宇宙射线	10^{-14}

在强烈的科学求知欲驱使下，贝克勒尔说服巴黎理工大学（Ecole Polytechnique）当时最杰出的年轻科学家之一玛丽·居里（Marie Curie），请她研究引起铀盐放射的究竟是什么物质。居里夫人的丈夫Pierre是一位在压电效应研究中颇有建树的物理

学家。夫妻两人一起全力以赴研究这种神奇的物质。经过研究，他们发现了一种比铀具有更强放射性的物质，且将其称为镭，并在 1898 年发表了这一发现。居里夫人后来用放射性活度（Radioactivity）来描述物质的放射特性。

就在居里夫妇研究放射性物质的同时，英国科学界也开始掀起一股研究原子结构的热潮。进入 19 世纪之后，道尔顿（John Dalton）的原子化学理论盛行，成为当时有关物质内部结构组成的权威理论。该理论认为所有物质都是由不可再分的原子组成的。但是，1897 年，汤姆生（J. J. Thompson）发现带有负电荷的电子比当时认为最轻的原子的质量还要小得多，于是，人们不得不重新建立有关物质的基本组成元素的新概念。如今我们所用的原子结构模型图就是卢瑟福（Ernest Rutherford）在 1911 年建立的。他证明，原子的主要质量集中于带正电荷的密实的原子核中，原子核周围环绕着带负电荷的电子云。后来，1913 年，波尔（Niels Bohr）提出了原子的行星理论，认为原子核周围环绕着许多质量非常小的沿轨道运行的电子。

根据这种基本原子结构理论，卢瑟福最先指出放射性辐射就是原子的自发衰变。这一基本思想受到广泛认同。卢瑟福于 1908 年荣获了诺贝尔化学奖，放射性物质的神秘面纱也自此开始被揭开。通过观察放射性辐射在磁场中的行为，居里夫妇发现放射性物质可以发出 3 种不同的射线，即卢瑟福命名的 α 射线、β 射线和 γ 射线。现在我们已经知道：α 粒子就是带正电的氦原子核；β 粒子是带负电的电子；而 γ 射线则是既没有质量也不带电荷的纯粹电磁辐射。

15.2.2　基本粒子

20 世纪 30 年代，人们已经清楚地认识到电子、质子和中子是构成原子的 3 种基本粒子。图 15.1 显示了这些粒子在原子中的分布情况。每个原子包含的粒子有：一

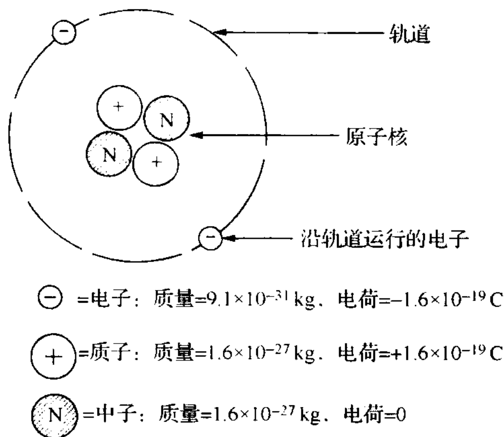


图 15.1　原子结构的行星图

注：原子的质量主要集中于原子核，原子核由质子和中子组成，并且带净正电荷，原子核周围环绕着沿轨道运行的质量很小的电子。稳定的原子中，沿轨道运行的电子所带的电荷总量与原子核相同，只是电荷极性相反。图中所示为氦原子的结构。

个或多个电子，每个电子的质量约为 9.1×10^{-31} kg，并带有 1.6×10^{-19} C 的负电荷；至少一个带正电的质子，质量为 1.6×10^{-27} kg，约为电子质量的 1800 倍；可能有中子，中子的质量与质子相同，但不带电荷。

由于单个原子核的质量非常小，如果用质量单位 g 来描述，就要用很大的负指数值。于是，人们专门定义了原子质量单位（atomic mass units，amu）来表示原子的质量，并用符号 u 表示。也就是，将碳 12（即 $^{12}_{6}C$）作为参考原子，把它的质量定为 12 u，每个 u 所代表的质量就等于碳 12 原子质量的 1/12。这样，用 u 这个单位就可以衡量其他原子的质量。最轻的氢原子同位素的质量约为 1 u。

元素的原子量以 g 为单位称为克原子量（gram-atomic mass）。物质的克原子量也称为摩尔质量。克原子量的大小正比于单个原子的实际质量，任何 1 mol 物质都包含一定数目的原子，1 克原子量物质中所包含的原子数是 6.023×10^{23}，这就是阿伏伽德罗（Avogadro）常数 N_A。

这样，以 g 为单位的任意一种原子的质量就等于其元素的克原子量除以阿伏伽德罗常数。例如，碳 12 原子的质量等于克原子量 12g 除以 N_A，即

$$^{12}\text{C 的质量}=\frac{12\text{g}}{6.023\times10^{23}}=1.99\times10^{-23}\text{g}$$

由于碳 12 原子的质量等于 12u，因此，1u 的质量就等于碳 12 原子质量的1/12，即

$$1\ \text{u}=\frac{1.99\times10^{-23}\ \text{g}/^{12}\text{C 原子}}{12\text{u}/^{12}\text{C 原子}}$$

$$1\ \text{u}=1.66\times10^{-24}\ \text{g}$$

与粒子具有质量一样，电子的电荷也是物质的基本性质。由于电子电荷是可能存在的最小电量，因此，通常将其表示为单位负电荷，即－1。如果用国际单位制（即 m、km、s）来表示单位电荷，那么其值为 1.6×10^{-19} C。这样，其他原子中所含粒子的带电量就都是这个值的整数倍。任何粒子所带的电量都不可能等于 2.5 倍电子电荷之类的值。

原子核中包含质子和中子，两者统称为核子。元素原子核所包含的核子总数称为原子量，也就是原子的质量数，用符号 A 表示。而单独质子数则称为原子序数，用符号 Z 表示。质量数相同的不同原子称为同量异位素（isobar），而中子数相同的不同原子则称为同中子异位素（isotone）。

在自然界，质子和中子可以有各种不同的组合。以氢、氘、氚这 3 种氢原子的同位素为例。氢的原子核只含有一个质子；重氢（即氘）的原子核由一个质子和一个中子组成；氚的原子核则包含一个质子和两个中子。另外，由两个质子和两个中子组成的稳定粒子（即 α 粒子）是自然界存在的氦原子的原子核。放射性物质发射的 α 射线就是这种粒子流组成的。

3 种氢原子的原子核含有相同数目的质子，即它们的原子序数 Z 相同，但中子数目不同，也就是具有不同的原子量 A。具备这种性质的原子互称为同位素，“同位素”一词来源于希腊语，意思是“位置相同”。同位素泛指任何原子，尤其是具有放射性的原子。

不过，现在人们喜欢用核素这个名称来表示质子和中子的特定组合。于是，同位素就是具有相同原子序数的核素。所有原子序数 Z 大于 83 或者原子量 A 大于 209 的元素都具有放射性，也就是，它们会自发衰变成其他元素，而衰变期间就会发射活性粒子。

原子有好几种符号表示法，阅读这方面的文献资料时，首先要熟悉这些符号表示法。美国的表示法是在化学元素符号的前面写上原子序数作为下标，并在后面写上原子量作为上标，如$_{53}I^{131}$。由于化学元素符号本身其实已经包含了原子序数的信息，因此，人们就常简写为 I^{131}。欧洲的表示法则习惯于将原子量作为上标写在元素符号的前面，如^{131}I。为了统一各种表示法，1964 年推出了国际标准，将原子量作为上标写在化学元素符号之前，即^{131}I。如果不用上标，那么就用"钴 60"之类的语言来表示。例如，氦原子的简写就是^{4}He，其中 4 表示原子量，是原子核中质子数与中子数之和。

15.2.3 原子结构与辐射

如前所述，根据波尔所提出的模型，原子结构类似于行星系统，其中的电子在环绕原子核的离散轨道（即电子壳层）上旋转运行，电子轨道的半径由主量子数 n 决定，$n=1$，2，3…，是一些整数值。对应于各个不同的 n 值，从小到大各个轨道依次冠以 K、L、M 和 N 等名称，并沿用至今。1925 年，泡利（Wolfgang Pauli）根据量子力学原理修正了波尔的原子结构模型。泡利的模型能够解释许多现象，也可以计算电子在不同轨道壳层上的结合能。他还发现原子中的单个电子可以用如下 4 个量子数来描述：主量子数 n，它是一个整数且为标量；角量子数 l，它是一个矢量，取值为 $0\sim n-1$ 之间的整数；磁量子数 m_l，取值为 $-l$ 至 $+l$ 之间的整数；自旋量子数 m_s，取值为 $\pm 1/2$。根据泡利不相容原理，一个原子中的任意两个电子都不可能具有一组完全相同的量子数 n、l、m_l 和 m_s。

对于呈电中性的原子，所有原子的轨道电子总数刚好等于原子核所带的正电荷数目。原子的化学性质由轨道电子决定，这些电子的活动形成了分子之间的键合、各种光谱、荧光和磷光发射等。另一方面，处于内层轨道上的电子受到的约束力较大，只有在辐射等相当大的能量作用下才能够离开轨道。

将电子驱出轨道所需的能量必须等于电子所处轨道层的结合能，处于最内层轨道上的电子结合能最大。如果将电子从内层轨道转移到外层轨道，那么所需的能量就等于两个轨道层的结合能之差。这个能量差的大小是化学元素具有的一种自然特性。当外层轨道的电子跃迁到内层时，如果原子将这种特征能量差以光子的形式释放出来，那么，就产生了 X 射线辐射；如果原子不发射光子，而是将能量转移给其他轨道上的电子，那么接收能量的电子就会被驱出所处的轨道，这种电子被称为 Auger 电子。电子跃迁产生 X 射线的几率被称为物质的荧光产率（fluorescent yield）。

质子与中子之间（即核子之间）受到一种短程束缚力的作用，称为核力，它比质子之间的电磁排斥力要大得多。与电子轨道相似，核子的运动通常也用壳层模型

来描述。不过，核子的运动形式很有限，它们由一组核量子数来定义。最稳定的核子分布形式称为基态，另外还有亚稳态和激发态这两种广泛存在的分布形式。其中，亚稳态下，核子处于一种不稳定状态，不过，在转变成其他状态之前，具有相对较长的寿命；激发态下，核子非常不稳定，瞬间就会转变成其他状态。因此，核子具有稳态和非稳态两种能级。具有相同原子序数 Z 和原子量 A 但能级不同的原子核被称为同质异能素（isomer）。非稳态核子最终总是要通过吸收或者释放能量（以光子或者粒子的形式）转变到稳定的基态，这个转变过程称为放射性衰变。如前所述，自然界存在的原子序数 Z 大于 83 的重元素都是不稳定的。

在判断核子的稳定性时需要计算核子的结合能。结合能反映了克服核力从而分离核子所需的最小能量，它可以通过基本质能公式 $E=mc^2$ 来计算，其中 E 为能量，m 为质量，c 为光速。我们常用电子伏特（eV）这个能量单位来表示各种质量，其根据就是该质能理论。原子质量往往小于其所包含的质子、中子和电子各个组分的质量之和。这个明显的质量损失 Δm 通常称为原子的质量亏损，它就是由核子的结合能引起的，等价于能量变化 Δmc^2。例如，如前所述，中性碳 12 原子的质量等于 12.0 u。但是，碳 12 原子各个组成成分的质量之和却等于 12.10223 u。两个质量之差 0.10223 u 就等价于核子的总结合能 95.23 MeV，再除以原子量 $A=12$，就得到每个核子的结合能 7.936 MeV。对于原子量大于 11 的原子核，每个核子的结合能在 7.4～8.8 MeV 范围之内。根据质能公式，质量为 1.6605655×10^{-24} g 的 1 个原子质量单位（即 1 u）的质量等价于 931.502 MeV 能量。另外，原子中电子的质量很小，每个只相当于 0.511 MeV 能量。

例 15.1 请证明 1 个原子质量单位（即 1 u）释放的能量约等于 931 MeV。

解： 应用爱因斯坦质能转换公式

$$E=mc^2$$

式中 m——质量，单位为 g；

c——光速，单位为 cm/s；

E——能量。

因此，1 个原子质量单位的等价能量为

$$E=(1.66\times10^{-24}\ \text{g})\times(3\times10^{10}\ \text{cm/s})^2$$

$$E=1.49\times10^{-3}\ \text{g}\cdot\text{cm}^2/\text{s}^2$$

在物理学中，单位 $\text{g}\cdot\text{cm}^2/\text{s}^2$ 用 erg[⊖] 表示，即

$$E=1.49\times10^{-3}\ \text{erg}$$

另一个常用能量单位是电子伏特（eV）。1 eV 等于 1 个电子在 1 V 电压差形成的电场中被加速时所获得的动能大小。电压差 V 作用于电荷 e 上时所做的功为 Ve，且 1 个电子所带的电荷等于 1.6×10^{-19} C，因此，1 电子伏特（eV）的能量即为

$$1\ \text{eV}=1.6\times10^{-19}\ \text{C}\times1\ \text{V}=1.6\times10^{-19}\ \text{J}$$

⊖ $1\text{erg}=10^{-7}\text{J}$。——译者注

也就是

$$1\ \text{eV}=1.6\times10^{-12}\ \text{erg}$$

由于电子伏特（eV）是一个很小的能量单位，因此，常用 keV 和 MeV 这两个单位，即

$$1\ \text{keV}=1000\text{eV}$$

$$1\ \text{MeV}=1000000\text{eV}$$

现在，可以用 MeV 来表示原子质量单位了，即

$$1\ \text{u}=1.49\times10^{-3}\ \text{erg}$$

且
$$1\ \text{MeV}=1.6\times10^{-6}\ \text{erg}$$

因此
$$1\ \text{u}=931\ \text{MeV}$$

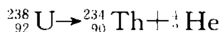
可见，如果将 1 u 质量全部转化成能量，其大小约为 931 MeV。1 MeV 的能量只能将 1 mg 重的物体举起 10^{-6} cm，由此可以体会其大小。■

非稳态核子常称为放射性核素，在经过一定的寿命期之后，最终它会通过发射电离辐射回到稳定状态，其寿命遵循某个特定的概率分布。一般地，放射性核素有天然和人造两种。天然放射性核素能够自发产生辐射，无需外界提供能源；而人造放射性核素是用高能量粒子轰击处于稳态的核子产生的。这两种放射性核素在发射扫描检测和核医学中都发挥着重要的作用。放射性核素的基本特性有平均寿命（即半衰期）、衰变模式、发射特性（即电离辐射的类型和能量）等。

无论是天然的还是人造的，任何放射性物质都按照几种相同的模式衰变，也就是发射 α、β、γ 等射线。此处所说的 α 和 β 射线都是由母元素分裂成子元素时产生的。放射性核素从非稳态转变到稳态时所经历的衰变模式可以分成如下 6 种：α 衰变（即 α 辐射）；β^- 衰变（即 β 负电子衰变）；β^+ 衰变（即 β 正电子衰变）；电子俘获（Electron Capture，EC）；同质异能跃迁（Isomeric Transition，IT）；裂变。下面分别介绍这些衰变模式。

1. α 衰变

α 粒子就是高速运动的电离的氦原子核 ${}^4_2\text{He}$。因此，原子核每释放 1 个 α 粒子，就要损失 2 个质子和 2 个中子，从而使其 Z 值减 2，而 A 值减 4。很多天然重元素具有这种衰变。例如，母原子核 ${}^{238}_{92}\text{U}$（铀元素）发射 1 个 α 粒子之后，就变成子原子核 ${}^{234}_{90}\text{Th}$，它是钍元素。此衰变过程可以表示为



$${}^{238}_{92}\text{U}\rightarrow{}^{234}_{90}\text{Th}+{}^{4}_{2}\text{He}$$

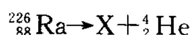
注意：根据电荷守恒定律，此反应式左边和右边的原子序数（即质子数）必须相同，也就是 92＝90＋2；反应式左边和右边的原子量（即质子数加中子数）必须相同，也就是 238＝234＋4。

α 衰变之类的某种元素转变成另一种元素的过程称为蜕变。要产生 α 辐射，母原子的质量必须大于子原子与 α 粒子的质量之和。在衰变过程中，多余的质量转变成能量，并表现为子原子核和 α 粒子的动能。由于 α 粒子的质量远小于子原子核，而衰变过程中动量必须守恒，较轻的 α 粒子反冲的速度要比子原子核大得多，因此，大

部分动能都被 α 粒子带走。一般地，核衰变过程中总是质量小的粒子带走大部分能量。

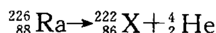
例 15.2 请说明镭元素$^{226}_{88}Ra$发生 α 衰变时会产生哪种子元素。

解：此衰变过程可以写为



$$^{226}_{88}Ra \rightarrow X + ^{4}_{2}He$$

其中 X 就是未知的子元素。由反应式两边原子量和原子序数相等可以求得，子原子核的原子量应该为 222，原子序数应该为 86，因此



$$^{226}_{88}Ra \rightarrow ^{222}_{86}X + ^{4}_{2}He$$

查元素周期表可知，原子序数为 86 的元素就是氡（Rn）。 ■

例 15.3 以上例 15.2 说明了镭元素$^{226}_{88}Ra$经过 α 衰变转变成氡元素$^{222}_{86}Rn$的过程，请计算此过程释放的能量。已知$^{226}_{88}Ra$的质量为 226.025406 u，$^{222}_{86}Rn$的质量为 222.017574 u，$^{4}_{2}He$的质量为 4.002603 u。

解：衰变之后，子原子核的质量 m_d与 α 粒子的质量 m_α之和为

$$m_d + m_\alpha = 222.017574\ \text{u} + 4.002603\ \text{u} = 226.020177\ \text{u}$$

将母原子核的质量记为 M_p，则衰变过程的质量损失为

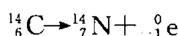
$$\Delta m = M_p - (m_d + m_\alpha) = 226.025406\ \text{u} - 226.020177\ \text{u} = 0.005229\ \text{u}$$

利用关系式 1 u=931.50 MeV，就可以求得所释放的能量为

$$E = 0.005229\ \text{u} \times 931.50\ \text{MeV/u} = 4.87\ \text{MeV}$$ ■

2. β⁻ 衰变与 γ 射线发射

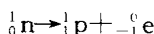
放射性核素发生 β^- 衰变时，子原子核的核子数与母原子核相同，而原子序数 Z 却增 1。如下是一个典型的 β^- 衰变：



$$^{14}_{6}C \rightarrow ^{14}_{7}N + ^{0}_{-1}e$$

这里，碳元素和氮元素符号的上标和下标与前述规则一致，但是，电子符号的上、下标需要解释一下。－1 表示电子所带的电荷大小与质子所带电荷相等，而极性为负。0 则作为电子的质量数，表示电子的质量与碳和氮的原子核相比，几乎为零。

通常认为原子核中只含有质子和中子，那么，它怎么会发射电子呢？这似乎很奇怪。其实是由于原子核的中子转变成质子的过程中产生了电子。此过程可以表示如下：



$$^{1}_{0}n \rightarrow ^{1}_{1}p + ^{0}_{-1}e$$

例 15.4 请计算$^{14}_{6}C$经过 β^- 衰变转变成$^{14}_{7}N$时所释放的能量。

解：$^{14}_{6}C$的质量为 14.003242 u，$^{14}_{7}N$的质量为 14.003074 u，因此，衰变前后的质量差为

$$\Delta m = 14.003242\ \text{u} - 14.003074\ \text{u} = 0.000168\ \text{u}$$

与其相对应的能量为

$$E = 0.000168\ \text{u} \times 931.50\ \text{MeV/u} = 0.156\ \text{MeV}$$ ■

实际上，只有少数电子可以达到这么大的动能。β^- 衰变时发射的多数电子的动

能都小于这个计算值。那么，如果子原子核和电子没有带走所释放的全部能量，奇怪的是，这些丢失的能量到哪里去了呢？能量必须守恒呀！

泡利于 1930 年指出，为了保持能量守恒，肯定存在第三种粒子，带走了“丢失”的能量。后来，费米（Enrico Fermi）把这种粒子命名为微中子，因为它必须呈电中性，且几乎没有静止质量。隐藏多年之后，微中子终于在 1950 年被检测到了。微中子用符号 v 表示，它具有如下特性：

1）不带电荷；

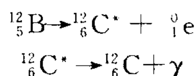
2）静止质量比电子还要小；

3）与物质之间的相互作用非常微弱，很难检测到。

磷 32 是一种典型的纯 β^- 发射物质，即 ^{32}P 转变为 ^{32}S，它已被用于临床治疗。

核素在放射衰变过程中经常会逗留在激发态能级上，然后，产生二次衰变，释放一个或多个光子，降到较低的能级状态。这种过程与原子发光很相似。当原子中的电子从高能态“跃迁”到低能态时，原子就会产生辐射，释放多余的能量。同样，原子核也就是通过类似的方式来释放衰变等核事件之后产生的多余能量。不同的是，在原子核去激发过程中，释放能量的“跃迁”发生于质子或中子上，是质子或中子从高能态移到低能态。这种去激发过程所发出的光子称为 γ 射线，它的能量比可见光大得多。一般经历 β^- 衰变的放射性核素几乎都同时发射 γ 射线。例如，碘 131 会经历数次 β^- 衰变并同时发射 γ 射线。

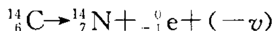
如下反应式表示了一个典型的包含 γ 射线发射的核衰变过程：



第一个反应式表示 ^{12}B 转变为 $^{12}C^*$ 的 β^- 衰变，其中星号用于表示碳 12 原子核处于激发态。然后，这种激发态的碳原子核通过发射 γ 射线到达基态。注意，γ 射线发射不会改变元素的 Z 值和 A 值。

3. β^+ 衰变与 γ 射线发射

引入微中子之后，β^+ 衰变过程就可以写成如下形式：



“$-v$”前面的短横表示这是反微中子，即与微中子对应的反粒子。为了解释什么是反微中子，我们来看如下衰变过程：



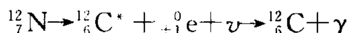
可见，当 ^{12}N 衰变成 ^{12}C 时，所产生的粒子与电子一样，只是带正电荷 $+e$。这种粒子称为正电子。除了电荷的极性以外，正电子的其他方面都与电子相同，这种正电子就称为电子的反粒子。

注意，β^+ 衰变的正电子产生于原子核，如同 1 个质子变成了 1 个中子加 1 个正电子一样。正电子衰变产生一种新元素，其 Z 值增 1，而 A 值不变。例如，碳 11 衰变为更稳定的同位素硼，即 ^{11}C 变成 ^{11}B。与负电子一样，正电子也与微中子分享能量。正电子衰变也伴随着 γ 射线的发射。但是，正电子一旦发射，常常会在 1 ns 时间之

内就迅速与电子碰撞而湮灭，并产生一对动量为 0.511 MeV 的反方向运动的光子。因此，正电子衰变所需的最小能量为 1.022 MeV。

例 15.5　^{12}N 经过 β^+ 衰变成为 ^{12}C 激发态，而后再衰变到基态，同时发射能量为 4.43 MeV 的 γ 射线。请问在此过程中所发射的 β 粒子的最大动能是多少？

解： 此 β^+ 衰变过程可以描述为



$$^{12}_{7}N \rightarrow ^{12}_{6}C^* + ^{0}_{+1}e + \upsilon \rightarrow ^{12}_{6}C + \gamma$$

由能量守恒定律可知，核素 X 通过发射 1 个或多个粒子（用 x 表示）衰变为较轻的核素 X^1 时，X 的质量总是大于 X^1 与 x 的质量之和，剩余的质量称为衰变的 Q 值。本例题的 Q 值可以表示为

$$Q=[m(^{12}_{7}N)-m(^{12}_{6}C^*)-2m_e]c^2$$

为了求得该衰变过程的 Q 值，首先要计算处于激发态的 ^{12}C 的质量。由于基态 ^{12}C 的质量为 12.000000u，因此，其激发态的质量为

$$12.000000u+\frac{4.43MeV}{931.5MeV/u}=12.004756u$$

于是 $Q=[12.018613u-12.004756u-2\times0.000549u]\times931.5MeV/u=11.89MeV$

这就是发射的 β 粒子可能具有的最大动能。■

4. 电子俘获（K 俘获）

通常，在内层轨道（即最接近原子核的 K 轨道）上运行的电子可以被原子核俘获，就像质子俘获电子转变成中子一样。电子俘获（EC）过程产生一种新元素，其原子序数 Z 减 1，而原子量 A 不变，类似于 β^+ 衰变。其实，某些放射性核素同时具有一定的正电子衰变几率和电子俘获几率。例如，铁 52 衰变时，电子俘获的几率约为 42%，β^+ 发射的几率约为 58%。并且，电子俘获过程可能伴随 γ 放射。例如，铬 51（^{51}Cr）转变为钒 51（^{51}V）时，有 90% 直接进入基态，剩余的 10% 则先到达 ^{51}V 的激发态，然后再通过发射光子进入基态。电子俘获造成的内层轨道空缺也会引起 X 射线发射或者 Auger 电子辐射。电子俘获过程不存在高能电子（即 β 粒子），因此，生物组织的辐射吸收较小。

5. 同质异能跃迁与内转换

处于亚稳态的放射性核素只发射 γ 射线，其元素的原子量 A 保持不变，是一种同质异能转变，其异构体的原子量记为 Am。例如，锝 99m（即 ^{99}mTc）衰变成 ^{99}Tc 等。不过，除了发射光子之外，核素的能量还会以一定的概率转移到内层轨道电子上，这就是所谓的内转换（Internal Conversion）。内转换电子几乎都是单能 β 粒子。例如，钡 135m 具有同质异能衰变，发射内转换电子的几率约为 84%，同时在内层轨道上产生电子空缺，导致 X 射线或者 Auger 电子辐射。

6. 原子核裂变

一般地，重核可以裂变成 2 个新核，并且或多或少会有相等的裂变碎片。这种裂变可以自发产生，但多数由中子轰击产生。例如，铀裂变产生的碎片的原子序数

大部分在 42 与 56 之间。氙 133 等很多医用放射性核素都是核裂变的产物，经过适当的放射化学处理，就可以将这些裂变产物提取出来。

15.2.4 放射性衰变

原子核经过分裂转变成另一种原子核的现象称为放射性衰变，它是所有放射性物质的一种特性。放射性物质的衰变速率是个常数，它不受温度、压力等变化以及化学反应的影响。此外，衰变过程是一个随机事件，放射性元素中的每个原子都具有相同的衰变概率。

随着衰变过程的进行，显然，能够分裂的原子核数目会越来越少。剩余原子核中单位时间内衰变的原子数目称为衰变常数 λ。半衰期与衰变常数相关，衰变常数越大，衰变速率就越快，半衰期也就越短。衰变过程中，衰变常数恒定不变。虽然各种放射性核素的分裂过程具有不同的衰变常数 λ 和半衰期 $T_{1/2}$，但所有原子核的衰变方式都相同，只是速率不同而已。衰变速率由放射性元素的特性决定。

假设某种放射性核素在 t 时刻的原子数目为 N，则衰变速率定义为 $-\mathrm{d}N/\mathrm{d}t$（负号表示减少），它与原子数目成正比，即

$$\mathrm{d}N/\mathrm{d}t=\lambda N \tag{15.1}$$

式中 λ——衰变常数。

假设 $t=t_0$ 时的初始原子数目为 N_0，将上式求积分可得

$$N=N_0\mathrm{e}^{-\lambda t} \tag{15.2}$$

式中 N——t 时刻放射性核素的原子数目，其曲线如图 15.2 所示；

e——自然对数的基数，即 e=2.718。

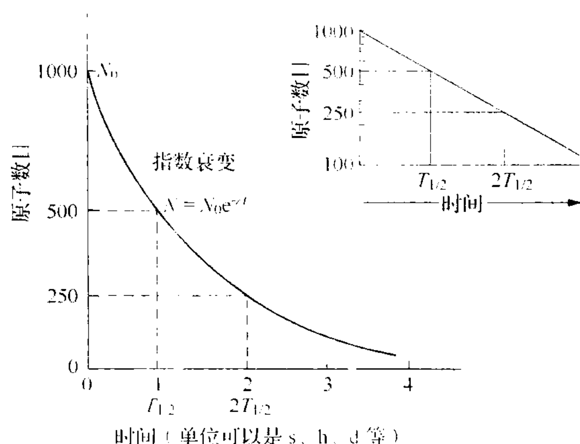


图 15.2 放射性物质呈指数形式衰变（在以自然对数为坐标的示图中（右上角），衰变曲线是一条直线）

居里 Ci 是放射性活度的单位，其定义为

$$1\text{Ci}=3.7\times10^{10}\text{ 核衰变/s}$$

它表示每秒核衰变数，约为 1g 镭的衰变活度，它是老的活度单位。如今新规定的放射性活度单位是国际单位制的贝可勒尔，记为 Bq，即

$$1\text{Bq}=1\text{ 核衰变/s}$$

因此，$1\text{Ci}=3.7\times10^{10}\text{Bq}$。最常用的活度单位是毫居里（即 10^{-3}Ci）和微居里（即 10^{-6}Ci）。

放射性物质最常用特性之一是半衰期 $T_{1/2}$。所有放射性物质都遵循同相的衰变模式，经过一定的时间间隔之后，原来物质中的一半核素发生衰变，然后再经过同样的时间间隔，剩余物质中的一半核素又衰变掉了，依此类推。

半衰期 $T_{1/2}$ 就是一半数目放射性核素衰变所需要的时间。令式（15.2）中的 $N=N_0/2$，就可以求得 $T_{1/2}$，即

$$N=N_0\text{e}^{-\lambda t}$$

所以

$$T_{1/2}=\ln2/\lambda=0.693/\lambda$$

例 15.6 请推导半衰期 $T_{1/2}$ 的计算公式。

解：由 $N=N_0\text{e}^{-\lambda t}$ 可知，经过半衰期 $T_{1/2}$ 的时间间隔之后，剩余的放射性核素数目为 $N=N_0/2$，因此

$$N_0/2 = N_0\text{e}^{-\lambda T_{1/2}}$$

此式两边同除以 N_0，可得

$$1/2 = \text{e}^{-\lambda T_{1/2}}$$

然后，两边同取自然对数，由于 $\ln \text{e}=1$，可得

$$\ln1/2=-\lambda T_{1/2}$$

$\ln1/2=-0.693$，于是

$$-0.693=-\lambda T_{1/2}$$

$$T_{1/2}=0.693/\lambda$$

■

假设 $t=0$ 时刻放射性核素的数目为 N_0，那么，经过第一个半衰期之后，剩余核素的数目为 $N_0/2$；再经过第二个半衰期之后，核素的数目又减少一半，成为 $N_0/4$；依此类推。各种不同物质的放射性半衰期的变化范围非常大，为 10^{-22}s～10^{21}年。

例 15.7 放射性核素 ^{286}Ra 的半衰期为 1.6×10^3年。如果某个样品含有 3×10^{16} 个这种核素，请计算其放射性活度。

解：首先计算衰变常数 λ。由于

$$T_{1/2}=1.6\times10^3\text{年}=(1.6\times10^3\text{年})\times(3.15\times10^7\text{s/年})=5.0\times10^{10}\text{s}$$

因此

$$\lambda=\frac{0.693}{T_{1/2}}=\frac{0.693}{5.0\times10^{10}\text{s}}=1.4\times10^{-11}\text{s}^{-1}$$

$t=0$ 时刻样品的放射性活度，即衰变速率 R_0，就可以用导数 $-\text{d}N/\text{d}t=N_0\lambda$ 求得，其

中 N_0为 $t=0$ 时刻放射性核素的数目。本题中 $N_0=3\times10^{16}$，于是

$$R_0=N_0\lambda=(1.4\times10^{-11}s^{-1})\times(3\times10^{16})=4.2\times10^5\text{核衰变/s}$$

由于 $1Ci=3.7\times10^{10}$核衰变/s，因此，样品的放射性活度（即衰变速率）就是

$$R_0=11.3\mu Ci$$

■

核医学的发展与放射性核素产品的开发密切相关，而放射性核素的选择主要取决于它与医学诊断和治疗相关的物理特性，以及是否可以将其结合到适用于生物医学研究的化合物中。

放射性药物是含有放射性核素标记的特殊化合物，它可以是简单的无机盐，也可以是复杂的有机分子。众所周知，化学元素的性质反映了它可能产生的生物学效应，周期表中位于同一组的几个化学元素都具有相似的生物学效应。例如，锶 85 已被用于标记骨骼中钙离子的代谢过程。表 15.2 归纳了目前常用于核医学的一些重要放射性核素。

表 15.2 核医学中常用的放射性核素

放射性核素	半 衰 期	衰 变	产 生 方 式	化 学 形 态
碳-11	20.38min	β^-	回旋加速器	C11-L-甲基-蛋氨酸
氟-18	109.77min	β^-	回旋加速器	氟代脱氧葡萄糖
磷-31	14.29 天	β^-	反应堆	磷酸盐
铬-51	27.704 天	EC	反应堆	铬酸钠
钴-57	270.9 天	EC	回旋加速器	维生素 B_{12}
镓-67	78.26h	EC	回旋加速器	柠檬酸化合物
钼-99	66.0h	β^-	反应堆	钼酸盐
锝-99m	6.02h	IT	发生器	TcO_4及其化合物
铟-111	2.83 天	EC	回旋加速器	二乙烯三胺五乙酸（DTPA）
碘-123	13.2h	EC	回旋加速器	碘化物
碘-123	60.14 天	EC	反应堆	蛋白质
碘-131	8.04 天	β^-	反应堆	碘化物
氙-133	5.245 天	β^-	反应堆	气体
铊-201	3.044 天	EC	回旋加速器	氯化铊

15.2.5 辐射量及其单位

至此大家已经清楚，所有放射性物质都会衰变，并在衰变过程中发出 α、β、γ 等各种辐射，这些不同的辐射通常混合在一起，很难单独检测。物质的放射性通常有如下 3 种测量方法：计数放射性物质在 1s 时间内的衰变数目；检测放射性辐射引起的带正电荷或负电荷的原子（或离子）的产生效率；测量辐射穿过某种物体时物体所吸收的能量。这 3 种基本测量原理分别对应有 3 种不同的辐射计量单位，即居里 Ci、伦琴 R 和辐射吸收剂量 rad。

注意，Ci 定义的是单位时间内的衰变数目，而不是 α、β、γ 等辐射类型，它只衡量放射源的活度。而 R 和 rad 则是反映辐射对于受照射物体的作用的计量单位。因

此，Ci定义辐射源，而R和rad则定义辐射源对物体的作用。X射线和γ射线的主要作用之一是使原子产生电离，也就是使原子变成带正电荷和负电荷的离子对。伦琴R的数值是指标准条件下（即温度为0℃且气压为760mm Hg㊀）1cm^{3}空气受到X射线（或γ射线）辐照所产生的离子对总数。由于离子对都带电荷，因此，可以通过测量相关的电效应来测定辐射。1R定义为能够使1kg空气电离产生2.58×10^{-4}C电量的离子对的X射线（或γ射线）剂量。可见，伦琴R是辐射剂量的单位，不是辐射强度的单位。辐射吸收剂量rad是指受照射物体所吸收的总能量。1rad等于能够使1kg受照射物质吸收0.01J能量的辐射量，这里，"J"就是国际单位制的能量单位焦耳。

核医学应用中人体组织可能受到各种不同放射性物质的辐射照射，相同的吸收剂量未必产生同样程度的生物效应，为了便于比较，常用人体伦琴当量雷姆（roentgen equivalent man，rem）这个单位来表示辐射的生物效应大小，它是一个反映生物组织所受到的一种或数种电离辐射的生物学剂量单位。

上述这些辐射量的单位都是核医学各种测量的常用单位。例如，伦琴R（较常用的是毫伦琴，即mR）是一般环境辐射测量仪的读数单位。rad则被用于描述注射了放射药物的人体脏器所受到的曝光剂量。雷姆rem用于表示胶片式辐射计量器等体检仪器的曝光剂量。

在美国核医学医用内照射剂量委员会（Medical Internal Radiation Dose，MIRD）的不断努力下，传统辐射剂量测定法正在被现代化的测定方法逐步取代。如今国际公认的生物组织辐射剂量单位是格雷（Gy），1Gy相当于能够使1kg组织吸收1J能量的辐射量。如前所述，辐射吸收剂量过去曾经用rad表示。1rad的能量相当于100erg/g，即0.01J/kg，等于0.01Gy。因此，1rad/mCi等于0.27027mGy/MBq，即1mGy/MBq等于3.7rad/mCi。不过，核医学中所用的辐射剂量测定法现在一般都用rad单位进行计算。

活体生物，特别是哺乳动物，在高剂量辐射的照射下，会被诱发产生病变，甚至会死亡。这种特性已被用于恶性组织病变的治疗。治疗时，将高剂量辐射传送到人体的某个局部区域，用以破坏病变组织。低剂量的辐射不会对人体产生明显的影响，但是，如果长期作用，就会增加诱发癌症的危险。如果人体的生殖器官受到辐射照射，还会增加后代产生遗传缺陷的概率。因此，必须尽可能降低相关工作人员的辐射曝光剂量，并制定有关核医学应用的指导方针。

美国核管理委员会（Nuclear Regulatory Commission，NRC）已经实施了各种标准，普通公众每年最大的辐射曝光剂量不能超过0.5Sv（即rem），职业人员的全身最大曝光剂量不能超过每3个月1.25Sv，肢体最大曝光剂量不能超过每3个月18.75Sv。因此，对工作人员和工作场所的监测是保证辐射安全的重要措施之一。工作人员的常规检查通常使用胶片式辐射剂量仪或者指环剂量计。环境辐射物的检查以及辐射物的清除必须定期进行，并且在辐射污染事件发生时进行及时处理。常

㊀ 1mm Hg=133.322Pa。

用的辐射屏蔽材料是铅，所需的厚度取决于 γ 射线的能量。另外，放射性废物必须存放在具有良好屏蔽的指定地方，这些废物所需采取的处理方法与其特性和半衰期有关，但都必须将它们放在临时存储装置中，并且最终都要按照规定程序进行处理。

15.3 检测仪器和成像设备

由上述辐射测量可知，辐射能量只能通过间接测量辐射所产生的效应来检测，几种间接测量方法如下：

摄像法：在 X 射线等照射下使胶片曝光、变黑，从而检测辐射量。注意，X 射线相当于 γ 射线，在本章后面有关外源性辐射测量中将讲述 X 射线。

电离法：当射线穿过检测器探头中的气体时会使气体电离，产生离子对，检测器通过收集并计算这些离子对来检测辐射量。这种方法用于 α 射线的测量最有效，而对 γ 射线的测量效果较差。

冷光法：冷光是指非灼热引起的发光。用穿透性较强的射线轰击某些物体时，物体会产生这类发光，通过检测这些闪光就可以计算出射线的强度。冷光法非常有用。下一节将讲述的闪烁探测器检测的就是电离辐射所引起的荧光效应。这种方法对 α、β、γ 这 3 种辐射都能很好地检测。

15.3.1 闪烁探测器

多数新型核医学检测系统都使用闪烁器探头。闪烁探测器几乎可以用于各种辐射射线的测量，只是对于不同的检测对象所用的闪烁器类型和结构有所不同。不过，无论何种应用，所有闪烁器探头的基本原理都是一样的。硫化锌和碘化钠等物体在电离辐射照射下都会发出闪光，因而被称为闪烁物。闪烁物发光的量与其所吸收的能量成正比，变化范围很大。如果将闪烁物置于光电倍增管等电子器件的感光表面上，那么，光电倍增管就可以把闪烁物所发出的光转化为一系列很小的电脉冲信号，这些脉冲的幅值正比于 γ 射线等入射光的能量。经过放大处理之后，用户就可以从这些电脉冲信号中获得有关照射闪烁物的射线的信息，包括辐射量等性质。这种闪烁器可用于测定放射性核素在病人体内某个或某些器官组织中的含量和分布，用于医学诊断。

图 15.3 显示了闪烁探测器的基本结构，它包括：由闪烁晶体、光电倍增管和前置放大器组成的检测部分；由线性放大器和单通道脉冲幅值分析器等组成的信号处理部分；由计数器、扫描仪和示波器等组成的数据显示部分。

当闪烁晶体检测到辐射事件时，光电倍增管电路就会产生电脉冲，但其产生的脉冲信号非常弱。为了避免噪声引起的信号严重失真，并且能够获得足够强的信号用于后续处理和显示，就要用放大器增加脉冲幅值，这就是线性放大器。

由于轰击闪烁晶体的各种不同 γ 射线的能量变动范围很大，因此，闪烁检测器

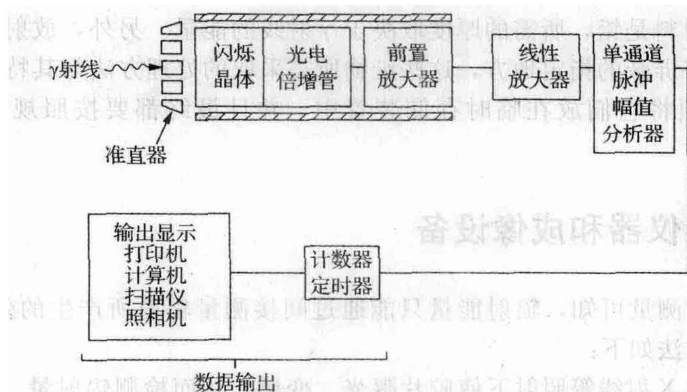


图 15.3 闪烁探测器的基本结构示意图

的线性放大器所接收到的电脉冲幅值的变化也很大。如果用户需要检测的是探头所接收到的各种辐射能量的总和，那么，只要用线性放大器就足够了。但是，如果用户只想检测其中某种特定放射性核素的活度，那么，就需要再加上一个脉冲幅值分析器。利用这个部件，用户就可以提取出与某种特定放射性核素相关的电脉冲信号，从总辐射能量中区分出某种辐射的能量。核医学实验室通常配备有各种单通道和多通道脉冲幅值分析器。

然后，这些测量数据必须显示出来，并且能够提供人体内部放射性活度及其所处位置的信息，这些信息对于各种人体脏器状态的诊断十分重要。医生既可以通过分析病人体内放射核素活度随时间的变化过程，来确定各种脏器的功能是否正常，也可以通过分析放射性核素的位置来显示脏器和异常组织的形态，这两方面的分析结果都很有价值。

闪烁探测器用于各种非成像检测时，一般需要对所产生的闪烁事件进行计数，且通过某种方式显示测量结果。因此，计数器是核医学中实现这种功能的最常见电子器件。计数器计数闪烁探头所产生的电脉冲，得到的计数值就代表了辐射能量的大小。计数器通常有两种工作方式，一种是预先设定计数时间，另一种是预先设定计数值。在预设时间的工作方式下，计数器在某个设定时间间隔内计数所发生的事件，时间一到，就自动停止计数，此时得到的计数值就是要测定的变量。在预设计数值的工作方式下，当计数器数到所设定的计数值时，自动停止计数，此时计数时间是要测定的变量。

两种变量的数值都可以通过计数器面板读出，也可以输入到其他部件，作进一步处理。例如，在预设时间的模式下，测量结果有如下几种显示方式：连续显示每秒钟、每分钟等单位时间内的计数值；用数字打印机连续输出随时间变化的计数值数据列表；直接输入计算机，保存计数值，并且在数据采集过程中进行各种实时计算处理，并将计算结果以放射衰变曲线的形式显示在监视器屏幕上。总之，闪烁探测器探头检测到辐射射线之后，再经过计数器的处理，就可以通过多种显示方式给

用户提供具有临床价值的信息。

闪烁探测器探头输出的数据也可以用于成像。一般，成像设备从探测器电路中获得上述电脉冲输出信号，并将这些脉冲信号与体内各个放射源的原始空间位点对应起来。显然，这种成像方法存在失真，因为实际放射源是三维空间中的位点，而多数普通仪器只能显示二维图像，是三维信息在二维平面上的投影。不过，尽管存在这个明显的缺点，这种方法仍然表现出极大的临床应用价值。

最常用的稳态闪烁探测器检测系统就是 γ 照相机。这种装置可以连续拍摄辐射场的各个部分，操作很像照相机，能够快速成像。γ 照相机最早出现于 20 世纪 50 年代末，由 Hal Anger 发明，它是核物理学、电子学、光学和数据处理等多个学科与临床医学结合的成果，其发展和应用不仅对于核医学，而且对于整个临床诊断都具有深刻的影响。Anger 创立的 γ 照相机的基本原理已经成为特定生理过程成像技术的基础。

15.3.2 γ 照相机

图 15.4 显示了 γ 照相机的基本工作原理。检测器探头放在需要扫描的人体脏器上方。γ 射线不会弯曲，总是直线传播。为了将来自脏器特定位点上的辐射传送到检测器的相应位点上，同时将来自其他位点的无关辐射挡掉，使探头只获取沿某一投影线传过来的 γ 射线，闪烁晶体前面必须加上准直器。准直器用于吸收无关辐射，一般用重金属制作，上面有很多平行小孔。

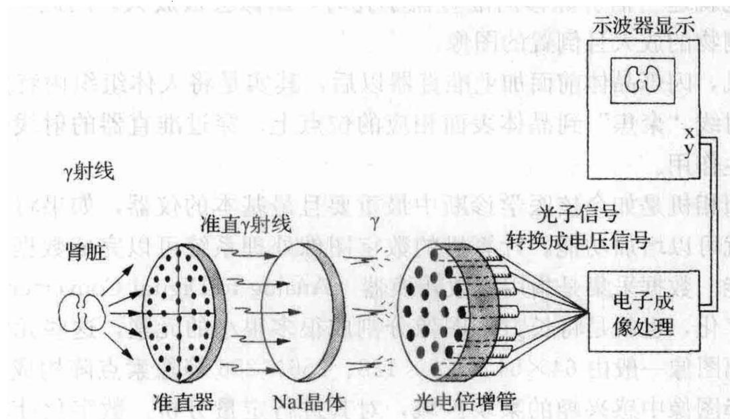


图 15.4 γ 照相机的基本工作原理

图 15.5a 所示就是一种多孔准直器，它是一块铅制平板，上面钻有许多小孔。可见，只有直接正对各个小孔发射的 γ 射线才能穿过准直器，并只照射到闪烁晶体表面的某个位点。远离准直器的位点（例如图中 y 区域）所发射的 γ 射线会投射到晶体表面多个位点，使检测分辨率大大下降。这里，分辨率定义为检测器区分与准直器相距不同距离的两个辐射源的能力。因此，对于多孔准直器而言，被检测区域越

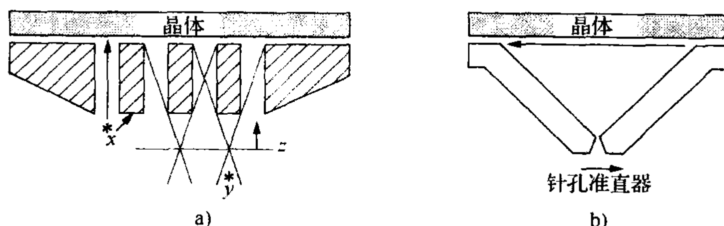


图 15.5 图 a 为与闪烁检测器结合使用的多孔准直器（图中 x 处发射的射线只在晶体的一个位点上成像，而 y 处发射的射线却可以到达晶体上多个位点）；图 b 为针孔准直器，它具有放大功能，用于近距离观察较小的器官组织

靠近准直器，分辨率就越高。如图所示，如果将被测物从 z 位置移向检测器，就可以提高分辨率。显然，在观察皮肤下的生物组织时、影像的分辨率与探头到达组织的距离相关。

除了这种多孔准直器以外，还有其他类型的准直器，图 15.5b 所示是只有一个小孔的针孔准直器。这种准直器只允许射向其小孔的射线通过。穿过准直器的 γ 射线沿直线传播，照射到晶体上，呈现出空间位置倒置的图像。如果辐射源到准直器之间的距离等于针孔到晶体之间的距离，那么，辐射源在晶体上成像的大小与其实际大小正好相同。适当调节被测物与准直器之间的距离，就可以放大或者缩小检测器的视野。也就是当辐射源移向准直器小孔时，图像会被放大。因此，针孔准直器可以呈现被测物的放大且倒置的图像。

由此可见，闪烁晶体前面加上准直器以后，其实是将人体组织内特定空间位点上所发出的射线“聚焦”到晶体表面相应的位点上。穿过准直器的射线撞击闪烁晶体并与其产生作用。

简单 γ 照相机是如今核医学诊断中最重要且最基本的仪器，如果将其与数字计算机相连，就可以增加功能。计算机的数字图像处理系统可以完成数据采集、存储和分析等功能。数据采集是指用模数转换器（Analog-to-Digital Converter，ADC）将图像信号数字化，也就是将长方形图像分割成很多很小的元素，这些元素称为像素(pixel)。每幅图像一般由 64×64、128×128、256×256 等像素点阵构成。这样，我们就可以选择图像中感兴趣的某块区域，对其进行定量分析。数字化计算还推进了动态图像研究的发展。如果从人体某个区域连续拍摄一系列图像，就可以容易地计算出该区域摄入和排出放射性核素的时间变化过程。例如，将这种方法用于心脏的成像时，就可以分析心室壁的搏动过程等。

15.3.3 正电子照相机

放射性核素衰变过程中所发射的正电子只能在生物组织中传播几毫米的距离，随后就会迅速湮灭，并产生一对以 180°相反方向传播的能量为 511keV 的光子。如果

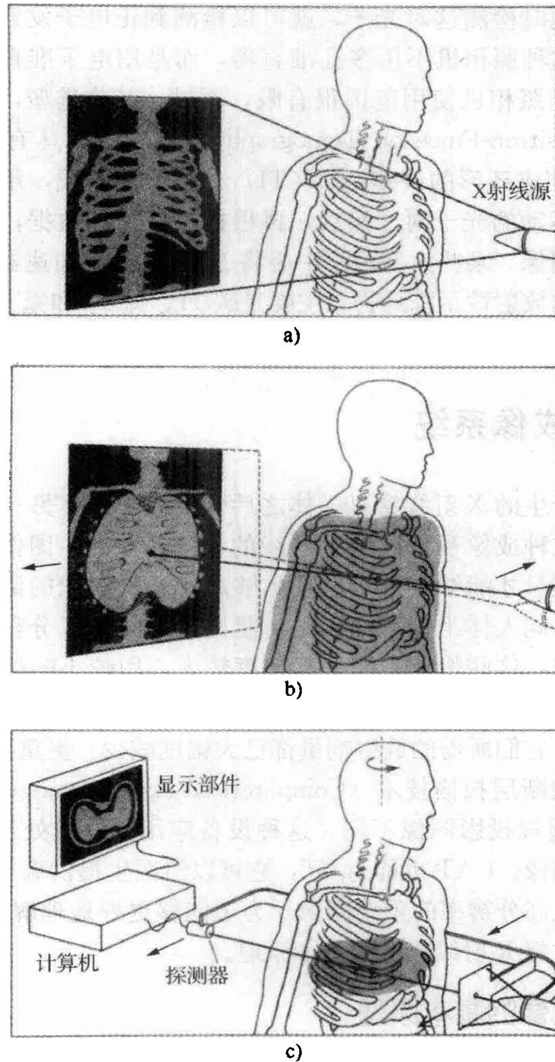
用两个闪烁探测器同时检测这对光子，就可以检测到正电子发射事件，这就是正电子照相机的原理。这种照相机不用多孔准直器，而是用电子准直的方法完成射线的选择。纯粹的正电子照相机使用范围很有限，不过，其改进型，即正电子发射断层扫描成像技术（Positron-Emission Tomography，PET）却具有非常重要的应用。PET的检测器由排列成环形的许多NaI（Tl）小检测器组成，用于检测正电子湮灭时产生的180°反向运动的光子对。然后，再根据这些检测数据，用计算机辅助图像重建技术建立断层图像。某些大型医疗中心将PET与回旋加速器配套使用，由加速器产生超短半衰期的放射性示踪药物（主要是碳11、氮13和氟18等）。这样，就可以同时获得被测对象的解剖结构和功能代谢两方面的信息。

15.4 X射线成像系统

本节介绍体外产生的X射线穿过人体之后，由置于人体另一侧的探测器检测并用于成像的技术。这种成像系统根据X射线的衰减系数产生图像。早期的设备需要使用相当大的放射剂量才能够产生对比度足够高的生物组织的清晰图像。对比度是指密度相差很小的不同人体组织（如脂肪与肌肉）之间的区分程度。软组织与肺部空气或者骨骼与肌肉，这些物体之间的对比度较大，用较小的放射剂量就能够区分出来。不过，随着胶片制作技术和新型探测器的不断发展，无论是高对比度还是低对比度的组织成像，它们所需的放射剂量都已大幅度减少。更重要的是20世纪70年代出现了计算机轴向断层扫描技术（Computerized Axial Tomography，CAT）。与常规X射线机产生的透视投影图像不同，这种设备应用计算机处理X射线衰减数据，可以重建人体断层图像。CAT也称为CT，它可以给医生提供病人任何部位任意断层上的具有高对比度且高分辨率的重建图像。为了能够更好地理解CAT成像技术，下面我们先回顾一下常规X射线成像的基本原理。

15.4.1 X射线成像的基本原理

任何X射线成像系统都由X射线源、准直器和X射线探测器组成。医疗诊断用的X射线系统一般使用能量为20～150keV的体外X射线源。自20世纪开始，常规X射线成像设备都用同样的方式成像，也就是，使用广谱X射线光束和照相胶片成像。一般，X射线由阴极射线管产生，这种射线管在高电压激励下会产生X射线光束。然后，X光束经过准直器校正之后，穿过人体，在成像平面上形成图像。不同的X射线系统使用不同的射线检测部件，如X光胶片、图像增强器、X光检测阵列等。如图15.6a～c所示，在标准X光胶片拍摄过程中，穿过人体的X射线将骨骼、器官、空腔和体内异物等投射到一张胶片上。这种“透视图”所呈现的其实是X射线穿过具有不同密度的人体组织和体液之后所发生的强度变化。



a)

b)

c)

图 15.6 3种X射线医学成像技术的原理

注：图a为常规X射线成像，从X射线源发出的光穿过人体之后，投射到一张胶片上；图b为断层成像，在胶片曝光过程中，X射线源向一个方向移动，而胶片则向另一个方向移动，这样，人体内只有某个垂直切面相对移动胶片保持不动，在投影图像中，该切面图像清晰，而其他切面图像都是模糊的；图c为计算机断层成像，将X射线源和探测器装在一个U形支架的两端，让支架水平移动并按照一系列不同的角度绕人体转动，使X射线多次扫描穿过人体某个横截面。然后，用一种特殊的计算机算法处理探测器记录到的X射线衰减数据，重建图像，并在显示屏上显示出来。

常规X射线成像技术的优点是分辨率高，对比度高，人体的曝光剂量较小，并产生永久记录图像。其缺点是几何失真明显，无法提供深度信息，也不能实时成像。因此，这种成像技术一般用于牙齿、胸部和骨骼成像。骨骼对于X射线的吸收很强，

所以用常规 X 射线成像很容易看出骨折。但是，这种成像是将三维物体投影成为二维平面图像，各种人体结构会在胶片上重叠，很难区分密度差别不大的人体组织。正是由于这个原因，常规 X 射线技术不能提供清晰可辨的颅脑图像，因为脑主要由软组织构成，密度差别较小。为了克服这些缺点，有人提出从许多不同角度拍摄透视图，这样，每张图像中人体内部各个器官之间的相对关系不同，可以用于辨别不同组织。另一种方法是使用对 X 射线要么透明要么不透明的造影剂（如空气或者碘溶液），用以增强图像的对比度。不过，这些方法都既费时，又难以实施，有时还存在较大的危险性，并且也不够精确。

20 世纪 20 年代初，人们开发了一种三维结构的 X 射线成像技术，称为平面断层成像，它可以将人体内某个垂直平面或横截面成像。如图 15.6b 所示，平面断层成像时，X 射线源沿一个方向移动，而放在人体另一侧的 X 射线感光胶片同时向另一个方向移动。这样，X 射线穿过人体的路径在不断变化。不过，每条 X 线在体内某个垂直平面（或横截面）上总是穿过同一个点。因此，在胶片上，这个感兴趣平面上的结构就被锐化，而其他所有平面上的结构则被模糊。显然，与传统 X 射线摄影相比，在用二维图像表示三维信息时，这种方法能够更好地显示人体各种结构的位置和细节。但是，它也存在缺点。首先，其深度信息存在误差，图像并不是真正锁定在单个平面上。其次，医学诊断中通常需要对比度大的高质量图像，便于读片，但这种平面断层图像的对比度较小。此外，断层扫描 X 射线的剂量要高于常规 X 摄影。并且，由于所需曝光时间较长，人体的活动也会降低成像质量。

计算机断层扫描成像（Computerized Tomography，CT）是一种完全不同的方法，其设备的主要结构包括扫描检测系统、计算机和显示部件。CT 将图像重建技术与 X 射线吸收测量相结合，可以显示出人体内部任何器官的 xy 二维轴向断层图，也可以沿 z 轴重建三维图像。如图 15.6c 所示，CT 的检测部分很像常规 X 射线机。经过准直器的一束 X 射线在某个平面上穿过被扫描的人体，到达放在人体另一侧的探测器。准直的细光束与探测器对准，每次只检测光束经过路径上的信息。X 光束穿过人体时，一部分能量被吸收，剩余的能量被探测器接收。探测器中通常装有碘化铯或钨酸镉等晶体，这些晶体受到 X 射线轰击时能够发射光子。这些光子的强度由探测器中的光电转换器检测，然后，再由计算机根据检测数据计算出人体组织吸收或者透射的 X 射线能量，用于成像。

CT 系统的 X 射线源和探测器通常安装在同一个扫描机架上，两者可以一起平移或者绕被测物转动。早期 CT 系统检测的是机架在各个不同旋转位置下 X 射线横扫整个断层平面所得到的人体 X 射线吸收分布数据。也就是先在某个旋转角度下平移机架获得一组 X 射线吸收数据，然后，再将机架转过一个很小的角度，进行下一次横向平移扫描，得到另一组吸收数据，以此类推。如此测得的每条 X 射线吸收分布曲线（即投影值）其实只是一维数据，其宽度与人体宽度一样，但厚度只有一个断层那么厚。

等距分布的检测位置的数目确定了最终所显示图像中像素的多少。例如，如果

要产生 160×180 个像素点阵的图像，首先需要平移 160 个等距分布的位置，测得一组含有 160 个 X 射线吸收值的数据。但是，要记住，这组一维数据只包含一条 X 射线吸收分布曲线。要获得下一条分布曲线，扫描机架要绕人体转过一个角度，再在新的角度下进行 160 次线性平移，获得一组新的测量数据。如此重复，直到扫描机架完成 180°旋转为止，一个断层的扫描结束。这些投影数据采集完成之后，共获得 160×180=28800 个数据。根据这些数据就可以计算出对应于每个像素的 X 射线强度值，并由此重建人体头部或身体某个断层截面的图像。

CT 系统的计算机具有非常重要的作用，它不断采集上述检测过程所获得的测量数据，并存入内存。等到所有 X 射线吸收数据采集结束之后，计算机就利用图像重建软件包分析这些数据。图像重建算法基于一组数学方程，原来用于天文学。由于每次重建所需的计算量太大，在计算机发明之前图像重建技术并不常用。正是现代化计算机技术的发展使得这些重建技术发挥了作用。

为了按照 X 射线吸收数据创建图像，根据所需的显示尺寸，计算机首先要用假想的网格将扫描的断层截面分成很多均匀的小单元体，就像网球拍上网线构成的网格一样。如图 15.7 所示，由于人体的截面有厚度，每个小单元体其实表示一个长方形柱体的组织块，其高度就是断层的厚度，宽度由网格阵列的大小决定。这种三维组织块被称为体素（voxel)，而在二维图像中则称为“像素”。

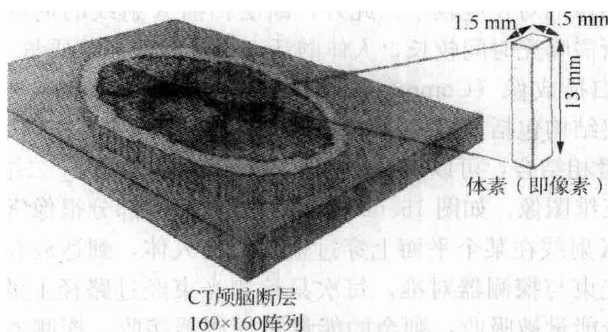


图 15.7 计算机断层成像技术原理示意图

注：每个断层被分成离散的三维组织块，称为体素。体素在计算机监视器上所显示的二维断层图像中被称为像素。图中所示的例子中，每个像素对应的实际生物组织大小为 1.5mm×1.5mm。

在上述 CT 扫描的例子中，每个体素要接受很细的一束 X 射线多达 180 次的照射，也就是在 180 个不同投影方向的测量数据中都包含了该体素的吸收值。由于每个体素在各次投影检测中的吸收值都一样，计算机就可以根据总吸收值和体素的大小，精确计算出该体素的平均吸收系数，并在二维图像中用相应像素点的灰度值表示出来。

例 15.8 请描述 X 射线束穿过人体断层截面时的衰减过程。

解：如图 15.8 所示，物体的断层截面可以看做由小方块构成，每个方块对于穿过的 X 射线都有衰减作用，会吸收入射光的部分能量。在图中所示的小方块中，假

设第 1 个方块对入射光的吸收系数为 A_1，第 2 个方块的吸收系数为 A_2，以此类推，第 n 个方块的吸收系数为 A_n，那么，这些方块对于穿过其中的入射光的总吸收系数 A 就等于各个方块的吸收系数之积。该总吸收系数的对数值就是吸收系数测量值。

再来看图 15.8，仅包含 4 个方块的物体的一组吸收系数方程如下：

扫描位置 1：$A(1)=A_1A_3$ (1)

扫描位置 2：$A(2)=A_2A_4$ (2)

扫描位置 3：$A(3)=A_1A_2$ (3)

扫描位置 4：$A(4)=A_3A_4$ (4)

实际上，这组方程中只有吸收系数 $A(1)$、$A(2)$、$A(3)$和 $A(4)$可以测得，它们是已知的，而 A_1、A_2、A_3和 A_4是要计算的未知数。4 个联立方程求解 4 个未知数，可以得到解。依此类推，如果要重建 n 行×n 列的断层截面图像，至少必须得到 n 个方向上的各 n 个测量值。例如，包含 320×320 像素的图像至少需要 320×320（即 102400）个独立的吸收系数测量值。 ■

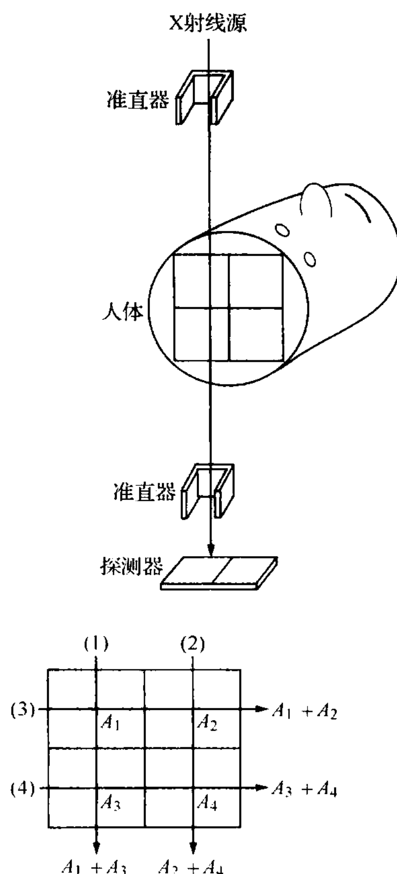


图 15.8 CT 图像每个像素吸收系数计算过程示意图

前面已经介绍过，CT 的吸收系数测量值是以一条条分布曲线的形式测得的。在各个扫描机架的旋转角度下，X 射线平移扫描的是与光束平行的某个平面（即断层），在此平面内，如果沿着垂直于 X 射线扫描束的方向，画出各条 X 射线的吸收测量值，就得到了吸收分布曲线，它表示每束 X 射线的总吸收量。一般，采样的分布曲线越多，所得到的图像分辨率就越高。

利用这些吸收数据，就可以重建二维平面的图像。如果将多个平面图像按照正确的顺序排列起来，还可以构成完整的三维图像。因此，三维图像的重建其实是从二维断层图像得到的，而二维图像又是从许多一维投影数据获得。实现这种计算的最早的方法称为反投影法，意思是将测得的分布曲线反向投影回到所测平面中的各个小区域。这种方法相当简单，但是它会产生模糊，因此该方法并不是很成功。后来，人们采用迭代法修正了反投影的分布曲线，获得了令人满意的图像。迭代法是个不错的方法，不过计算时间较长，一般需要经过好几次迭代计算，才能将原始分布曲线转化为可以反投影产生清晰图像的一组分布曲线。为了提高算法的速度，人们又开发了卷积滤波等数学方法将原始分布曲线直接转化为最终所需的分布曲线。

无论使用什么重建算法，最终的计算结果都是一样的，都要在计算机内存中产

生一个图形文件。该文件包含了图像中每个像素的吸收系数（即密度测量值），例如，160×160 图像就有 25600 个像素，而 320×320 图像的像素则有十万多个。图像每个像素的最终吸收系数计算值可以用图形的不同灰度和颜色表示出来。多数 CT 用计算机监视器屏幕显示图像。实际上，图形文件的像素数值反映的是被测人体断层截面上某个体积块的密度，确切地说是相对吸收系数。

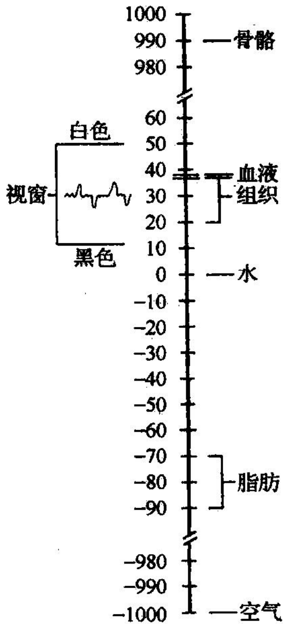
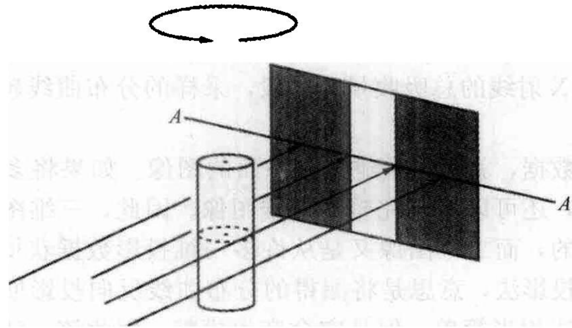


图 15.9 对应于吸收系数的 Hounsfield 标度

图 15.9 所示是 Hounsfield 定义的表示各种物质 X 射线吸收系数相对大小的标度，即 CT 值，其取值范围为 −1000～1000。CT 值最小的是空气（−1000），最大的是骨骼，水的 CT 值为 0，居于中间。整个 CT 值范围以水为中心，水的上下各分成 1000 个等级。这种设定是为了方便使用，因为生物组织的含水量很大，其吸收系数与水很接近。使用这种标度显示图像时，将图像每个像素的吸收系数都加上 1000，这样，空气的吸收值就变成了 0，而水的吸收值成为 1000。每个像素都有了标度值之后，就可以显示出重建图像了。

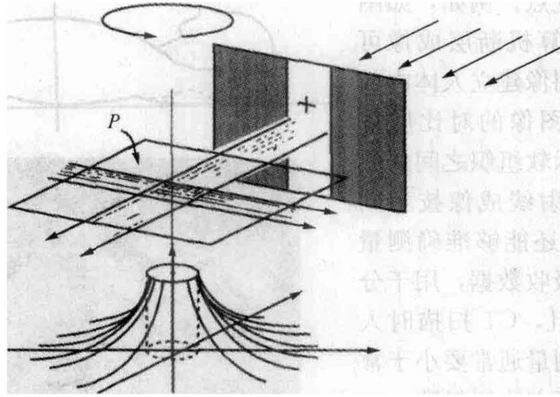
例 15.9 请阐述反投影处理方法。

解：反投影法可以通过以下简化模型来说明。如下图所示，假设有一束平行 X 射线直接穿过一个柱形吸收物，并在 X 射线胶片上留下投影。曝光后的胶片经过冲洗之后，胶片上沿如图所示的 *AA* 直线上显示出来的影像密度分布就是物体的投影值。如果围绕圆柱体，按照等间隔角度摄取一系列这种 X 射线影像，那么，就可以根据所获得的一组二维投影值重建 X 射线所穿过的断层截面的图像。



下图所示的反投影法是一种图像重建的近似算法。它将所测得的投影值依次按其原摄影路径反方向投影回去，并在各个相应的影像位置上进行叠加，从而获得扫描断层的重建图像。这种简单反投影法的缺点是会产生图像边缘失锐现象，即在边缘上出现伪像（如下图所示）。为了消除这种伪像，必须在反投影数据中减去原始影像两边的数值。在数学上，这种方法称为卷积法，相当于将每个投影

值减去某个与周围值成正比的数值。这是实际 CT 成像系统中使用的投影值修正方法。



■

对于任何医学成像设备所显示的图像，都必须弄清楚影响图像分辨率和精确度的因素，并加以修正。CT 扫描机采集的数据实际上反映了 X 射线在人体某个断层上的吸收分布，因此，仪器能够检测并投影的 X 射线的最小吸收变化量就决定了图像的密度分辨率。另外，图像上能够分辨出来的两个不同物体之间的最小距离称为空间分辨率。密度分辨率反映仪器检测生物组织之间差别的灵敏度，而空间分辨率则表示仪器所生成图像中可能分辨的细节。两种分辨率之间存在一定的联系，并且都与人体所接收的放射剂量相关。图像的清晰度反映了所测得的吸收数据的精确度，它常受到低信噪比的影响。如果射线穿过人体后到达探测器的光子数量较少，那么，信噪比就较低，相邻像素吸收系数之间的差别就减小，因而图像锐度降低。使用低放射剂量成像时都会遇到这个问题。不过，CT 系统的显示处理可以使重建图像得到增强，这是 CT 系统的一个重要功能。

CT 的显示部件通常能够同时显示给定数目（如 16 种、32 种）颜色级别或者灰度级别，这些显示级别经设定可以用于表示某个特定范围的吸收系数。这样，图形的吸收系数取值范围就可以用整个可用的灰度或者颜色级别范围来表示。为了提高分辨率，突出图像的细节，可以将黑白两色之间的灰度级范围限定在一个很小的 CT 标度值范围之内。设置这种可调的“视窗”就可以增强人体特定器官的图像。例如，如图 15.9 所示，如果要观察心脏组织的图像，就可以把整个视窗的窗位设置在水的 CT 值上方；如果要观察肺部组织的细节，那么，就可以把视窗下移到某个合适的位置。减小视窗的窗宽，可以增加 CT 系统的总灵敏度，从而进一步增强区分特定器官及其周围组织吸收系数的能力。

改善图像分辨性能的方法还有很多。例如，可以将图像的某一部分放大，或者将某些吸收值增亮。不过，这些方法其实并不增加原始信息量，只是通过处理计算机内存中存储的数据来增强图像。由于人肉眼观察图像时只能分辨与周围环境之间存在足够对比度的物体，因此，图像增强方法都是改变能够提高观察人员工作效率

的图像特性，使得图像中的物体便于定位并精确测量。

与常规X射线技术相比，CT扫描具有很多独特的优点。例如，如图15.10所示，CT计算机断层成像可以通过一系列二维图像建立人体内部结构的三维图像。图像的对比度很高，可以清晰地显示软组织之间的微小差别，而常规X射线成像技术则对此无能为力。CT还能够准确测量各种组织的X射线吸收数据，用于分析组织的特性，而且，CT扫描时人体所吸收的X射线剂量通常要小于常规X射线断层成像等方法的剂量。

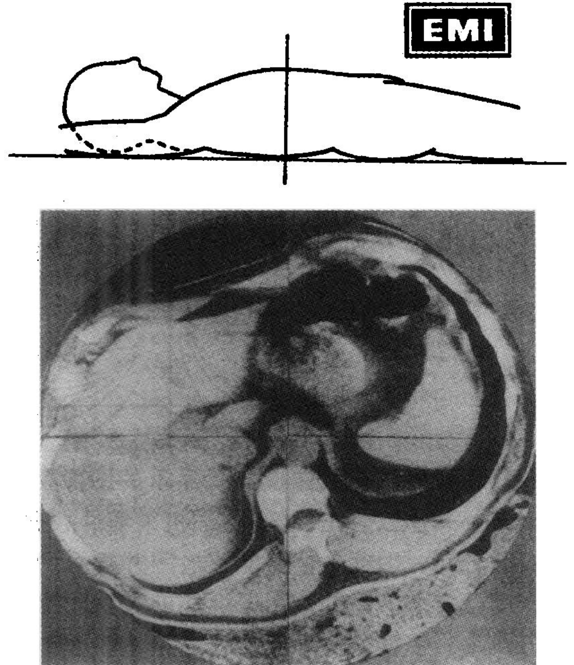


图15.10 人体CT断层图像

如图15.6所示，常规断层成像技术也可以通过将成像平面之外的物体模糊掉的方法，显示出断层图像，不过，这种方法只利用了某个断层上的射线扫描信息，断层两侧其余大部分穿过人体的射线信息都变成了图像中无用的伪像。而CT成像中的X射线经过整个被测断层平面，因此，所进行的测量100%与所测断层相关，并且只与所测断层相关，不受上下其他层面上物体的影响。所以，CT充分利用了全部X射线的测量信息，图像就比较清晰。

CT成像在医学中的应用非常广泛，各种软肿瘤、硬肿瘤、囊肿、血块、损伤或坏死组织等人体异常形态结构都能容易地从CT图像中分辨出来。CT系统具备的各种软组织区分功能极大地简化了很多医学诊断过程，使得其他某些需要住院、费用高的方法，或者对病人存在潜在危险的侵入性方法都被淘汰了。由于CT可以准确定位被测人体组织的各个区域，并监测实际治疗效果，因此，它不仅用于医学诊断，还用于放射治疗。

15.4.2 CT成像技术

CT扫描机问世以来一直在不断完善和发展，已经成为全世界举足轻重的一项技术。高质量CT系统的制造需要具备X射线技术、物理学、结晶学、电子学和数据处理等方面的坚实基础。

CT扫描机的发展已经取得了以下几个方面的重要成果：提高了X射线管的性能；缩短了扫描时间；提高了图像重建速度；减少甚至消除了图像中的伪像；减少了病人所受的放射剂量。这些成就主要涉及两个关键问题：第一，图像质量及其与医学诊断价值之间的关系；第二，仪器效率及其与病人保健和经济效益之间的关系。为了更好

地了解现代化新型 CT 系统的工作原理和功能，下面我们回顾一下 CT 的发展历史。

CT 扫描机通常由三大部分组成：扫描机架，它按照一定的扫描方式采集图像重建所需的数据；数据处理部分，它将扫描采集的数据转换成清晰的图形信息，再将这些信息显示成视觉图像，并提供各种增强图像的辅助功能，能够协助医生作出医学诊断；存储部件，它能够使用户在扫描结束后的任何时间随意复查扫描结果。

1. 扫描机架

CT 扫描机架的功能就是采集重建断层图像所需的信息。所有采集的信息都必须有用且准确，还必须有足够的相互独立的数据用于重建图像，使图像具有足够高的空间分辨率和密度分辨率，便于医生进行准确的诊断。可见，扫描系统的功能极其重要。

如图 15.11 所示，CT 扫描机的机架已经经历了数次重要的变革。第一代机架采用“平移一旋转”的扫描方式。其射线管和检测器安装在同一个机架上，通过机架的平移，让单束 X 射线扫描断层，从而获得分布在扫描方向上的一条吸收测量值曲线。每次平移扫描结束之后，机架转过 1°角度，再重复一遍平移扫描。直至转过 180°角度，完成总共 180 次平移扫描，并测得 180 条吸收分布曲线为止。整个“平移一旋转”扫描过程大约需要 4～5min。这种机架获得的图像质量很高，具有很高的临床诊断价值。但是，这种办法有几个缺点。首先，扫描时间太长，处理病人的速度

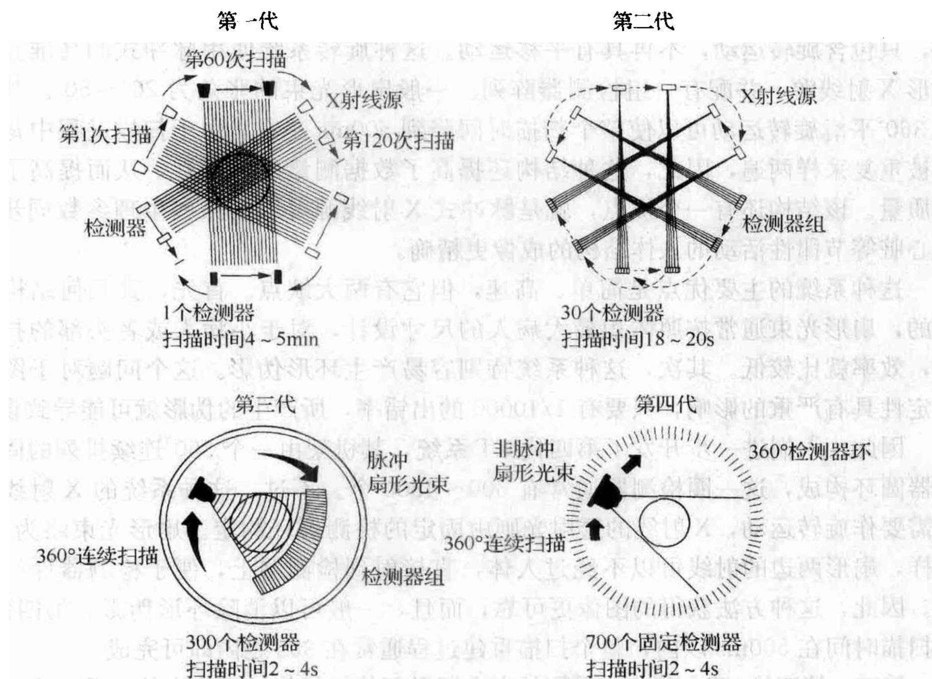


图 15.11　4 代 CT 扫描机架的不同结构

注：利用现代化的滑环技术，第三代和第四代 CT 可以实现断层厚度为 1～10mm 的旋转式容积扫描，像素可达 1024×1024。一般，病人一次屏气期间即可完成 50cm 长度的容积扫描。

相当慢，病人的周转量很小。其次，图像中常会出现条纹伪影，虽然对诊断一般没有什么影响，但图像质量确实有所降低。最后，最大的缺点可能就是扫描过程中人体必须保持不动，因此，这种CT扫描方式只能用于颅脑成像。

为了解决上述问题，人们开发了第二代CT扫描机架。这种机架也具有与第一代机架一样的平移扫描过程，但它绕人体180°的旋转步长从1°增加到了10°。同时，为了能够在每次平移扫描期间获得足够的吸收分布数据，这种机架使用了张角为10°的窄扇形光束。这样，机架每旋转一次就可以采集到间隔为1°的10条吸收分布曲线，再旋转10°，又可以采集下一组10条曲线。因此，一共只需旋转18次，进行18遍平移扫描，就可以获得整组180条吸收分布曲线。如果一遍平移扫描的时间为1s，那么，这种机架扫描一个断层约需18～20s。可见，第二代CT的扫描时间大幅度缩短，而整个180°扫描过程中所采集数据的精确度和数量并不受影响，利用这些数据就可以重建被测断层的图像。有了这种新型CT扫描机，就可以获得人体任何部位的断层图像。但是，这个扫描速度还是没有达到能够消除环形条纹伪影的速度，并且其速度不可能再提高了。其实，第一代CT和第二代CT扫描的速度都分别达到了其结构所限定的极限值，如果要确保一定的图像质量，扫描时间都不可能再缩短。不过，这两种设计结构可以提供的空间分辨率和密度分辨率是最高的。

随着CT系统各种部件的不断改进，出现了第三代CT系统，其机架有了重大改进，只包含旋转运动，不再具有平移运动。这种旋转系统使用脉冲式的高准直广角扇形X射线源，并配有一组检测器阵列。一般扇形光束的张角为20°～50°。其单纯的360°平滑旋转运动可以使整个扫描时间降到500ms左右。由于扫描过程中每个数据被重复采样两遍，因此，这种结构还提高了数据测量的可靠性，从而提高了成像的质量。该结构还有一个优点，就是脉冲式X射线源可以与某些生理参数同步，使得心脏等节律性活动的人体结构的成像更精确。

这种系统的主要优点是简单、高速，但它有两大缺点。首先，其几何结构是固定的，扇形光束通常按照体积最大病人的尺寸设计，对于小病人或者头部的扫描成像，效率就比较低。其次，这种系统特别容易产生环形伪影。这个问题对于图像的稳定性具有严重的影响，只要有1/10000的出错率，所产生的伪影就可能导致误诊。

因此，人们进一步开发了第四代CT系统，其机架由一个360°连续排列的固定检测器圆环构成，这一圈检测器通常有600～4800个。不过，这种系统的X射线源还是需要作旋转运动，X射线的透射光则由固定的检测器环测定。扇形光束略为增宽，这样，扇形两边的射线可以不经过人体，直接射到检测器上，便于检测器性能的校准。因此，这种方法获得的图像更可靠，而且，一般可以消除环形伪影。第四代CT的扫描时间在500ms以内，整个扫描重建过程通常在30s之内即可完成。

第三、第四代CT扫描机中采用的高电压滑环使机架能够围绕人体朝着一个方向连续旋转，此时，如果以一定的速度推进病床，穿过扫描平面，那么，X射线光束就会按照一个螺旋式盘旋的轨迹扫描整个人体。断层厚度一般为1～10mm，像素矩阵达1024×1024，病人一次屏气期间大约可以扫描50cm长度的容积。随着X射线管技术的发展，在

空间分辨率和密度分辨率（即对比度分辨率）保持不变的情况下，新的显像技术已经能够实现实时三维成像，并呈现动脉或大肠等管道结构中的物体流动的情况。

如图 15.12 所示，还有一种 CT 扫描机架结构，也许可以称为第五代 CT。它采用两组 210°弧形排列的固定检测器阵列，安装在病床的上方。另外，排列成 210°弧形的 4 个 X 射线靶环则安装在病床下方，与上方的检测器阵列相对。除了病床进出成像平面的机械运动以外，这种系统没有其他机械运动。

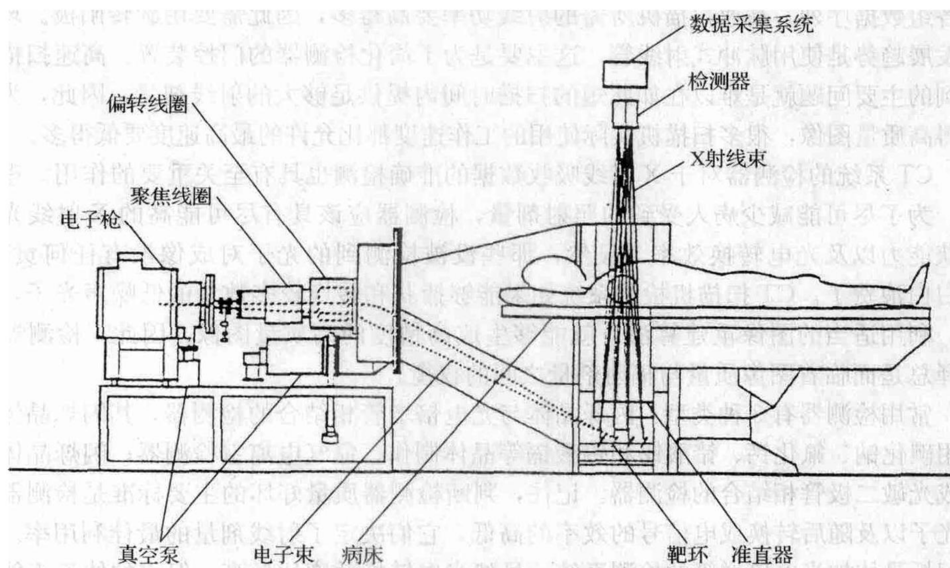


图 15.12　超高速 CT 扫描机示意图

注：其中，两个电磁线圈分别用于电子束的聚焦和偏转，使电子束能够穿过真空管，打到弧形靶环的各个位置上。4 个靶环产生的 4 束独立的 X 射线经过相应的准直器之后，射向人体。穿过人体的透射光由并排的两个检测器阵列检测，并产生 8 个独立的断层图像（此图由 Imatrol 馈赠）。

该系统的工作原理是：电子枪发射的电子束经过高压线圈所产生的磁场的聚焦和偏转之后，沿 210°弧形扫过 4 个 X 射线阳极靶环，靶环在高速电子束的轰击下，产生 4 束独立的 X 射线。X 射线经过准直器之后，穿过人体，并被并排的两个检测器阵列检测。这种系统在 50ms 或者 100ms 时间之内就可以同时完成多达 4 个 1.5mm 厚的断层扫描，图像的像素可达 512×512，因此，可用于心脏检查，诊断早期冠状动脉硬化等疾病。虽然这种超高速 CT 扫描机几乎不受人体脏器活动的影响，但目前它的对比度和空间分辨率都不如第三、第四代机械式 CT 扫描机，因此，其应用范围相对较小。

2. X 射线管和检测器

必须强调，CT 扫描检测的目标就是采集数以千计的人体 X 射线吸收系数的精确测量数据，这些在各个不同角度下测得的全方位数据包含了被扫描人体断层结构的大量信息。由于这些测量数据都是经过光子计数的方式获得的，因此，光子数量越多，数据的质量就越高。由此可见，只有射线剂量足够高时才能获得高质量图像。由于总辐射剂量由 X 射线管的最大功率及其工作时间决定，要达到一定的辐射剂量，

工作时间越短，所需功率越高；因此，高速扫描机必须使用高功率射线管。在设计X射线管和检测器时，射线剂量是必须考虑的重要因素之一。

目前，工作时间约为20s的CT系统都使用油冷式固定阳极射线管，这类射线管在扫描期间连续发射X射线，功率有限；而那些工作时间远小于20s的CT系统则使用水冷式旋转阳极射线管（或空气冷却射线管），这类射线管通常发射脉冲式射线，功率较高。固定阳极射线管工作时连续发射X射线，检测器则通过电子门控开关来采集各组数据序列。高速扫描机所需的射线功率要高得多，因此需要用旋转阳极。现在的发展趋势是使用脉冲式射线管，这主要是为了简化检测器的门控装置。高速扫描机遇到的主要问题就是难以在如此短的扫描时间内提供足够大的射线剂量，因此，为了获得高质量图像，很多扫描机实际使用的工作速度都比允许的最高速度要低得多。

CT系统的检测器对于X射线吸收数据的准确检测也具有至关重要的作用。理论上，为了尽可能减少病人受到的照射剂量，检测器应该具有尽可能高的X射线光子俘获能力以及光电转换效率。显然，那些没被检测到的光子对成像没有任何贡献，被白白浪费了。CT扫描机检测系统如果能够捕获和转换较多数量的低噪声光子，那么，利用适当的图像重建算法，就能够生成高精度的高质量图像。因此，检测器的选择总是面临着图像质量与辐射剂量之间的权衡。

常用检测器有3种类型：闪烁晶体与光电倍增管相结合的检测器，其闪烁晶体可以用碘化钠、氟化钙、锗酸铋和钨酸镉等晶体制作；氙气电离室检测器；闪烁晶体与集成光敏二极管相结合的检测器。记住，判断检测器质量好坏的主要标准是检测器捕获光子以及随后转换成电信号的效率的高低，它们决定了射线剂量的最佳利用率。采用闪烁晶体加光电倍增管的检测系统，虽然光电转换效率比较高，但它的体积不能做得很小，结构不紧凑，达不到捕获最大数量光子的目标。这种检测器的射线剂量利用率小于50%，因此，病人需要接受较高的照射剂量，才能弥补图像质量上的损失。氙气检测器虽然比较便宜、紧凑，适用于组建大型检测器阵列，但其光电转换效率较低，并且容易产生漂移。因此，使用氙气检测器也必须增加病人的照射剂量。

闪烁晶体与集成光敏二极管组合的结构可以排列得很紧密，其光电转换效率超过98%，射线剂量的利用率可达80%光子能量，图像质量很好，检测器性能也很稳定，因此，这种固态探测器在未来CT系统中的应用将越来越广泛。

3. 数据处理系统

除了上述扫描系统和检测系统以外，CT扫描机还需要一个快速数据处理系统，能够尽可能迅速地显示出扫描图像。随着扫描系统和检测系统的不断发展，CT的数据处理系统也结合最新的计算机技术在不断更新。例如，当第三代CT扫描机诞生时，其数据处理设备就必须增大采集和存储的数据量，这些数据包括：位置数据，如扫描机架所处的扫描位置等；参考检测器所测得的来自参考光的数据，用于监测X射线源输出的稳定性；每次扫描结束时采集的校准数据；检测器测得的大量X射线实际吸收数据。这些数据经过各种电子接口的转换之后，变成数字形式，被存入计算机内存。下一步就是图像重建处理（见图15.13）。

注意，图像重建的第一步是分析所有原始数据，并将它们转换成一组吸收系数分布曲线，一般有 180 条以上分布曲线。然后，再将这些分布曲线数据转换成可以显示的图像数据。图像重建是 CT 扫描机的核心部分，也是 CT 与常规 X 射线成像技术和许多其他成像技术截然不同之处。从一系列 X 射线吸收数据中重建生物组织 X 射线吸收特性的算法（即计算机程序）主要有以下几种：简单反投影法（也称总和法）、傅里叶变换法、积分方程法和级数展开法等。算法的选择主要取决于速度和精度。

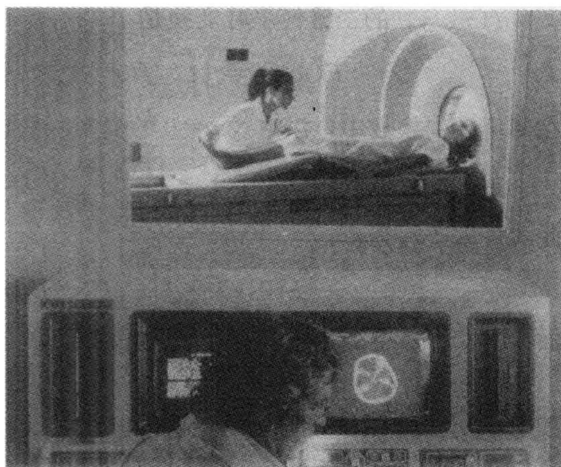


图 15.13 CT 扫描机显示终端的照片

图像重建中用到的数学计算方法有迭代法和解析法两种。迭代法也称代数重建法（Algebraic Reconstruction Technique，ART），Hounsfield 在第一代 CT 扫描机中使用的就是这种算法。该算法需要预先设定一组二维 X 射线吸收系数的初始估计值，然后用这组估计值计算出投影衰减数据的预测值，并将其与实际测量数据进行比较。然后，再用计算预测值与测量数据之差来反复迭代、修正初始估计值，直到计算得到的预测值与测量数据之差趋于 0 为止。一般，需要经过多次迭代之后，这种算法才能收敛。实际计算时，通常只要计算值与测量数据之差小于某个设定的误差限之后，就停止计算。ART 算法有很多版本，它们曾用于第一、二代 CT 扫描机。由于迭代法计算速度较慢，并且在存在噪声的情况下会出现无法收敛的问题，因此，新型 CT 扫描机都使用解析算法实现图像重建。

解析法包括傅里叶变换法、简单反投影法、滤波反投影法和卷积反投影法等方法。各种解析法都直接从衰减投影数据中重建图像，这是它与迭代法的根本区别。解析法使用中心切片定理和二维傅里叶变换。图 15.14 所示是中心切片定理示意图，假设图像函数为 $f(x, y)$，其沿 x 方向的一次投影所产生的投影曲线为 $g(y)$，那么

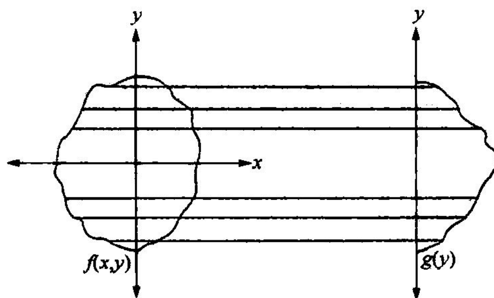


图 15.14 中心切片定理投影法示意图

$$g(y) = \int_{-\infty}^{\infty} f(x,y)\mathrm{d}x$$

该投影其实是如图 15.14 所示的一组线积分。这里，$f(x, y)$ 其实就是需要求解的重建图像，而 $g(y)$ 则对应于某个方向上 CT 扫描测得的 X 射线衰减投影数据。

$f(x, y)$ 的二维傅里叶变换是

$$F(u,v)=\iint_{-\infty}^{\infty} f(x,y)\exp[-\mathrm{j}2\pi(ux+vy)]\mathrm{d}x\mathrm{d}y$$

在傅里叶变换域中，沿直线 $u=0$ 的变换函数是：

$$F(0,v)=\iint_{-\infty}^{\infty} f(x,y)\exp(-\mathrm{j}2\pi vy)\mathrm{d}x\mathrm{d}y$$

此式可以改写为：

$$F(0,v)=\iint_{-\infty}^{\infty} f(x,y)\mathrm{d}x/\exp(-\mathrm{j}2\pi vy)\mathrm{d}y$$

即

$$F(0, v)=F_1[g(y)]_1$$

其中 F_1 [] 表示一维傅里叶变换。由此可见，$f(x, y)$ 在某个方向上投影的傅里叶变换在 $F(u, v)$ 域中是一条径向直线。因此，如果将各个不同角度下采集的投影数据进行傅里叶变换，就可以计算得到傅里叶变换域中的整个 $F(u, v)$ 函数。然后，再将 $F(u, v)$ 进行傅里叶反变换就可以求得 $f(x, y)$，从而获得重建的图像。

先进的 CT 扫描机使用滤波反投影图像重建算法。这种方法先将各个不同角度下测得的 X 射线衰减投影数据与一个空间滤波函数进行卷积计算，卷积可以在傅里叶域中进行，也可以直接在空间域进行。然后，如图 15.15 所示，再将滤波过的数据沿原投影路径反投影回去。各个不同角度的反投影数值之和就构成了最终的反投影数据。如果滤波是在空间域直接进行的，那么就可以直接得到反投影图像；如果滤波是在傅里叶域进行的，那么，经过傅里叶反变换就可以得到反投影图像。

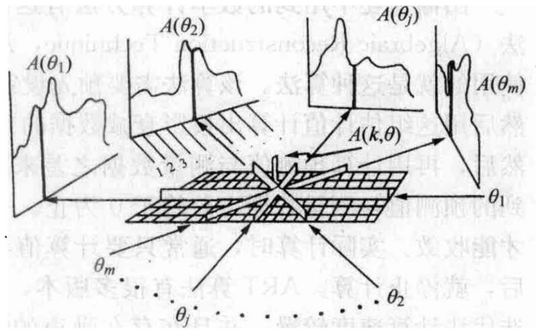


图 15.15 简单反投影算法示意图

总之，与早期的颅脑 CT 和躯体 CT 扫描机相比，如今的 CT 扫描机数据处理速度更快，X 射线能量的利用率更高，对于各种生物组织之间密度差异的检测灵敏度更高，空间分辨率也更高，并且成像中的伪影大为减少。各种 CT 扫描机性能的比较可以考虑以下 10 个因素：

1）机架的设计：它会影响扫描速度，病人处理时间和成本效益。

2）机架孔径的尺寸：它与病床的承重能力一起决定了能够接收的最大病人体积。

3）X 射线源的类型：它影响病人所受到的照射剂量和扫描设备的使用寿命。

4）X 射线扇形光束的张角和扫描范围：它们影响图像分辨率。

5）扫描断层的厚度以及射线源发射的脉冲数目和射线源的旋转角度：这些都是决定分辨率的重要因素。

6）检测器数量和类型：它们是影响图像质量的关键因素。

7）所用小型计算机的类型：它是决定系统性能和灵活性的重要因素之一。

8）系统配备的数据处理程序：它与用户的要求和系统可靠性密切相关。

9）系统的存储容量：它是便于数据存取的重要因素。

10）系统升级和通信：新型 CT 扫描机应该具备系统模块升级功能，并且能够连接网络进行通信。

在评价现有的 CT 系统时，不仅需要考察其各项性能指标，而且还要对照理想 CT 扫描机应该达到的要求。CT 扫描成像技术的最终目标是提供准确的诊断信息，促进人类医疗保健水平的提高。理想 CT 扫描机应该能够提供明确的诊断信息，使病人不需要再做其他测试就可以得到诊断结果。另外，经济成本也是必须考虑的重要因素。由于 CT 扫描机所需投资很大，维护费用也相当高，因此，理想 CT 系统应该能够快速接纳大量病人，具有较高的“病人周转量”（patient throughput）。也就是好的 CT 扫描机必须满足两个主要标准：能够提供高质量的诊断信息；有较高的病人周转量。

习　题

1. 请用符号形式表示例 15.2 的衰变过程。

2. 请计算 $^{238}_{92}U$（原子量＝238.050786u）通过发射 α 射线衰变为 $^{234}_{90}Th$（原子量＝232.038054u）时所释放的能量。

3. 碳 14（即 $^{14}_{6}C$）是碳的一种放射性同位素，其半衰期为 5730 年，如果初始样品中含有 1000 个 ^{14}C 核素，那么，22920 年之后还会留下多少个这种核素？

4. 假设某骨盆骨骼标本中提取的 50g 碳样品的 ^{14}C 衰变率为 200 核衰变/min，已知活体生物组织中碳的衰变率为 15 核衰变/min · g，并且 ^{14}C 的半衰期为 5730 年＝3.01×10^{9} min。请计算该骨骼标本的年龄。

5. 某放射性样本的半衰期为 30min，如果初始样品含核素 3×10^{16} 个，10min 之后还剩下多少核素？

6. 请计算 $^{14}_{6}C$ 转变为 $^{14}_{7}N$ 的 β 衰变过程中所释放的能量。

7. 半衰期为 140 天的某种钋样品衰变到原始强度的 1/10 需要多少时间？

8. 假设某种纯放射性物质的原始质量为 10^{3} g，2h 之后只剩下 0.25×10^{3} g，那么该物质的半衰期为多少？

9. 磷同位素的半衰期为 14 天，如果样品中含有 3×10^{16} 个这种核素，请计算其放射性活度。

10. 请计算放射性活度为 0.2μCi，半衰期为 8.1 天的样品中含有多少个放射性原子？

11. 某种新鲜制备的同位素样品的放射性活度为 10mCi，4h 之后其活度变为 8mCi

（a）请计算该同位素的衰变常数和半衰期；

（b）新鲜制备的样品中含有多少同位素原子？

（c）制备后 30h 其活度为多少？

12. 氚的半衰期为 12.33 年，5 年之后氚样品核素的百分之几已被衰变？

13. 衰变$^{23}_{10}Ne_{13} \rightarrow ^{23}_{11}Ne_{12} + ^{0}_{-1}e + v$所发射光子的最大动能为多少?

14. ^{235}U的半衰期为7.04×10^8年，假设一块4.55×10^9年之前凝固的岩石标本中含有N个^{235}U原子，那么，在它凝固时含有多少个^{235}U原子?

15. α、β和γ这3种基本辐射中哪一种对生物组织的穿透力最强，为什么?

16. 请阐述闪烁现象的机制。如何检测闪烁现象? 又如何能将其用于成像?

17. 治疗甲状腺癌的一种方法是将小剂量放射性物质直接注入肿瘤中，利用其发射的射线破坏癌细胞。请解释为什么$^{131}_{53}I$可以用于这种治疗?

18. 请举例说明临床上如何应用放射性物质检查人体器官的功能。

19. 假设某个光电倍增管中共有7个倍增极，电压分别为100V、200V、300V、400V、500V、600V和700V，且电子从倍增极表面发射所需的能量为10eV，那么，对于阴极产生的每个光电子（设其初始速度为0），第一级倍增极和最后一级倍增极分别会发射多少个二次电子?

20. 请阐述计算机断层成像的过程，并用图示加以说明。

21. 请描述迭代法重建图像的计算过程。

22. 请作图说明如何检测物体中各个体素的纯吸收系数。

推荐阅读资料

Alpen, E.L. (1990). *Radiation Biophysics*. Prentice Hall, Princeton, NJ.

Bronzino, J.D. (1982). *Computer Applications in Patient Care*. Addison-Wesley, Reading, MA.

Cho, Z.H., Jones, J.P. and Singh, M. (1993). *Foundations of Medical Imaging*. John Wiley, New York.

Croft, B.F. and Tsui, B.M.N. (1995). Nuclear medicine. In *The Biomedical Engineering Handbook*. CRC, Boca Raton, FL.

Cunningham, I.A. and Juoy, P.F. (1995). Computerized tomography. In *The Biomedical Engineering Handbook*. CRC, Boca Raton, FL.

Hounsfield, G.N. (1973). Computerized transverse axial scanning (tomography) part 1, description of system. *Br. J. Radiol.* **46,** 1016–1022.

McCollough, C.H. and Morin, R.L. (1994). The technical design and performance of ultrafast computed tomography. *Cardiac Imagery* **32**, 521–536.

McCollough, C.H. (1992). Acceptance testing of a fifth-generation scanning-electron-beam computed tomography scanner. *Med. Phys.* **19**, 846.

Serway, R.A. and Faughn, J.S. (1989). *College Physics,* 2nd Ed. Saunders College, Philadelphia.

Shroy Jr., R.E., Van Lipel, M.S. and Yaffe, M.J. (1995). *X-ray In the Biomedical Engineering Handbook*. CRC, Boca Raton, FL.

第 16 章 医学成像

Thomas Szabo 博士[⊖]

本章目录

⊖ Kirk K. Shung 和 Steven Wright 也参与了部分撰写工作。——原文注

本章主要内容：

- 脉冲回波测距与超声成像的区别；
- 超声图像的形成过程；
- 声波传播、反射和折射的基础知识；
- 压电换能器的特性和工作原理；
- 声阻抗匹配层的原理；
- 声波散射的基本类型；
- 声波吸收作用的计算；
- 声束形成和聚焦的基本原理；
- 超声成像系统的框图；
- 超声多普勒的基本原理；
- 磁场与电荷之间相互作用的4种情况；
- 原子核磁偶极矩在磁场中的自旋状态和进动；
- 拉莫尔频率与核磁共振；
- 翻转角对磁共振恢复过程以及各个弛豫时间常数的影响；
- 自由感应衰减信号与自旋回波信号的区别；
- 磁共振测量信号的空间定位；
- 创建磁共振图像的基本步骤；
- 磁共振图像所显示的信息；
- 磁共振成像系统的框图；
- 磁共振成像中 k 空间的作用；
- 磁共振成像应用中的功能磁共振；
- 几种主要医学成像技术的比较。

16.1 绪论

超声是最常用的重要医学成像设备之一，仅次于常规X射线摄影。近几年来，超声检查的累积数量估计已有几十亿。与X射线和CT计算机断层成像技术不同，超声成像没有电离辐射，因此，它对人体无伤害。并且，超声诊断仪便于携带，使用方便，成本低，且可以用于实时观察人体软组织和血流的动力学特性。超声成像检测的是来自生物组织的声波脉冲反射回波。

磁共振成像（Magnetic Resonance Imaging，MRI）也可以获得详细的人体解剖结构信息，它也没有电离辐射。MRI通过检测原子的自旋信号来区分置于静态大磁场中的人体内部的不同组织器官，显示人体骨骼和软组织等的静态断层截面图像。其成像结果与特定参数的选择有很大关系。在主要医学成像设备中，MRI最抽象，技术最复杂。它能够进行精确的解剖结构成像，因此，常被用于手术前的方案制定和癌症

检测。功能磁共振成像（即 fMRI）还可以显示大脑对于各种刺激的响应活动。

本章将讲述超声成像和 MRI 的基本原理及其应用。最后还将比较 CT、X 射线机等几种主要医学成像设备的主要特性。另外，本书第 10 章已经介绍过傅里叶变换的基本原理，它有助于理解本章将介绍的成像原理，因此，下一节首先回顾傅里叶变换。

16.1.1 傅里叶变换简介

傅里叶变换（Fourier Transform，FT）可以简化有关超声和 MRI 成像基本原理的论述，因此，这里首先回顾一下 FT 的一些性质。本书第 10 章介绍的傅里叶变换有一个重要性质，就是时域的卷积与频域的乘积之间的等价性。还有一个概念，就是在时域和频域上可以分别用冲激响应函数和传递函数来描述滤波器等系统的响应。

冲激函数是一个广义函数，它有一个特殊性质，当冲激函数与其他某个函数相乘并求积分时，就等价于对该函数的采样，即

$$\int_{-\infty}^{\infty} \delta(t-t_0)g(t)\mathrm{d}t = g(t_0) \tag{16.1}$$

冲激函数的傅里叶变换是一个指数，即

$$H(\omega) = \int_{-\infty}^{\infty} \delta(t-t_0)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}t = \mathrm{e}^{-\mathrm{i}\omega t_0} \tag{16.2}$$

可见，时域的延迟等价于频域乘以一个指数延迟因子。如果冲激函数没有延迟，即 $t_0=0$，那么，$H(\omega)=1.0$，为常数。

式（16.2）可以推广为如下一般形式（其中$\mathfrak{F}$表示傅里叶变换）：

$$\mathfrak{F}[g(t-b)]=\mathrm{e}^{-\mathrm{i}2\pi bf}G(\omega) \tag{16.3}$$

如果再加上一个比例因子，那么，该式所表示的 FT 位移（即延迟）性质就更通用了，即

$$\mathfrak{F}[g(a(t-b))]=\frac{\mathrm{e}^{-\mathrm{i}\omega b}}{|a|}G(\omega/a) \tag{16.4a}$$

注意，傅里叶反变换也具有相似的性质（反变换用 F^{-1} 表示），即

$$\mathfrak{F}^{-1}[G(a(\omega-b))]=\frac{\mathrm{e}^{\mathrm{i}bt}}{|a|}g(t/a) \tag{16.4b}$$

最后，冲激函数还有一个重要性质，即

$$\delta(at)=\frac{1}{|a|}\delta(t) \tag{16.5}$$

例 16.1　请计算 $R(\omega)=\sin[3(\omega-\omega_0)t_1]$的傅里叶反变换。

解： $R(\omega)$ 的基函数是

$$G(\omega)=\sin(\omega t_1)=\frac{\mathrm{e}^{\mathrm{i}\omega t_1}-\mathrm{e}^{-\mathrm{i}\omega t_1}}{2i}$$

于是，由式（16.2）的傅里叶变换可得

$$\mathfrak{F}^{-1}[G(\omega)]=g(t)=(i/2)[\delta(t-t_1)-\delta(t+t_1)]$$

设式（16.4b）的 $a=3$ 且 $b=\omega_0$，可得

$$\mathfrak{I}^{-1}[R(\omega)]=(e^{i\omega_0 t}/3)(i/2)\left[\delta(\frac{t-t_1}{3})-\delta(\frac{t+t_1}{3})\right]$$

根据式（16.5），可知 $r(t)$ 为

$$r(t)=(e^{i\omega_0 t})(i/2)[\delta(t-t_1)-\delta(t+t_1)]$$ ■

16.2 医学超声成像

16.2.1 超声成像的发展历史

虽然人们早就知道蝙蝠可以利用声音的回波进行定位，但是，出人意料的是，直到1913年泰坦尼克号沉没之后，人类才开始使用超声定位技术。泰坦尼克号灾难发生数月之后，英国科学家 L. F. Richardson 发明了利用声音在水中或者空中进行冰山等物体的回波定位方法，并获得了专利。第一次世界大战后期，法国的 C. Chilowsky 和 P. Langevin 使用高功率电子发射器和压电换能器，发明了回波定位装置，用于潜水艇定位和回波测距。回波测距原理的应用比电磁波雷达的应用要晚很多。雷达就是无线电探测和测距（Radio Detection And Ranging，RADAR）的简称，雷达中的回波测距线以极坐标的形式显示在平面位置显示器上（Plan Position Indicator，PPI），形成径向圆形扫描显示线，它其实可以看做一种早期的脉冲回波成像技术。这种显示技术后来被水下研究人员发展成为声呐，也就是声波导航和测距的简称（Sound Navigation And Ranging，SONAR）。

这些技术在第二次世界大战期间得到了推广和应用，第二次世界大战结束之后，有些医生想把回波定位原理用于人体内部物体的检测。碰巧，在大量废弃的军用设备中，有人找到了一种可用于检测固体物质内部缺陷的超声波反射测试仪。如图 16.1 所示，这种仪器可以发出几微秒长的周期性短脉冲声波，脉冲之间的时间间隔约为 1 ms。

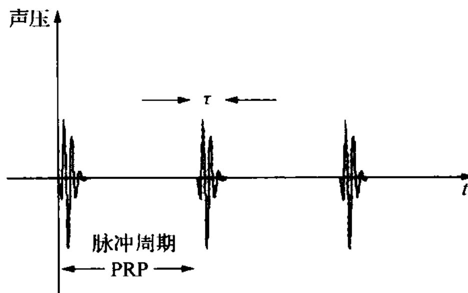


图 16.1 周期性发射的声波脉冲序列

注：脉冲重复周期称为 PRP（Pulse Repetition Period）。一般，为了辨别不同生物组织的边界位置，脉冲的 −3dB 宽度只有几微秒，而 PRP 约为 1ms。为了方便观察，图中的声波脉冲及其 PRP 并没有按比例画出。图中横坐标表示时间，纵坐标表示声压。对于疏密相间的纵波，稠密区域的声压为正值；稀疏区域的声压为负值。

生物组织边界位置的测定需要用短脉冲超声波。图 16.2 所示是一种典型的回波测距装置，图中下方显示了随时间变化的回波记录信号。其中，每个回波的时间延迟 t 等于声源与被测物体之间距离的两倍（记为 $2z$）除以声波在物质中的传播速度 c，也就是 $t=2z/c$。回波测距仪中有一个压电换能器，它既可以将电脉冲转换成声波脉冲，反过来，也可以将目标物反射回来的回波转换成电信号。回波电信号再由放大器放大之后，显示在示波器上，成为随时间变化的曲线。这种显示方式称为 A 型超声，其中 A 表示幅值。

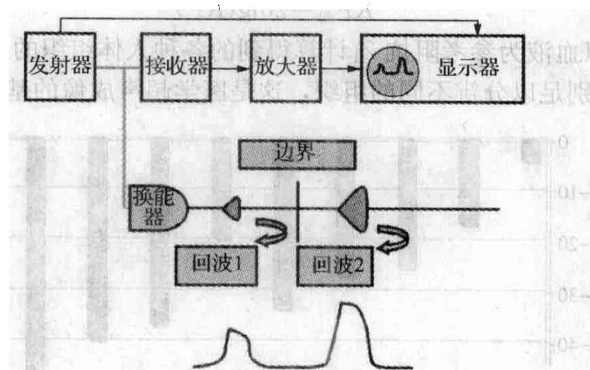


图 16.2 基本回波测距系统

注：主要包括发射器、换能器、接收器、放大器和显示器等部分，图中下方所示回波脉冲的不同延迟时间对应于被测反射物的不同深度。

1949 年，美国海军医生 G. Ludwig 首次发表了他测得的人体各个不同部位的声波传播速度，其平均值约为 $c_0=1540\mathrm{m/s}$。见表 16.1，后来的精确测量结果同样表明，多数生物组织的声速都很接近，差别很小，只有百分之几。这一重要发现意味着，只要用具有恒定声速 c_0 的简单延迟方程就可以计算人体不同脏器的位置。如今新型超声成像系统使用的仍然是这个原理。换言之，超声图像中人体脏器位置测定的精度很大程度上取决于测量范围内声波速度的均匀性。

表 16.1 生物组织的声特性

物　体	$c/(\mathrm{m/s})$	$\alpha/(\mathrm{dB/MHz^{y}\cdot cm})$	γ	$\rho/(\mathrm{kg/m^3})$	Z(M Rayl)
空气	343			1.21	0.0004
骨骼	3360	3.54	0.9	1789	6.00
血液	1550	0.14	1.21	1039	1.61
脂肪	1450	0.6	1.0	952	1.38
肾脏	2030			1420	2.89
肝脏	1570	0.45	1.05	1051	1.65
肌肉	1580	0.57	1.0	1041	1.645
20℃水	1482.3	2.17×10^{-3}	2.0	1000	1.482

Ludwig 还测量了生物组织的特征声阻抗，即 $Z=\rho c$，其中 ρ 和 c 分别是组织的密度和声速。他发现在两种不同生物组织的交界面处，声波的反射系数（Reflection

Factor, RF）与这两种生物组织的阻抗 Z_1 和 Z_2 有关，即

$$RF=\frac{Z_2-Z_1}{Z_2+Z_1} \tag{16.6a}$$

另一位美国医生 D. Howry 还发现，由于组织的反射系数很小，除了含有气体的区域和骨骼之外，超声很容易穿过多层软组织。1956 年，他利用声波作出了人体的详细解剖图，结果表明图中人体内部各种脏器和组织的位置和大小与已知数据相符。

声波反射系数的分贝值（即 dB）的计算公式为

$$RF_{dB}=20\lg(RF) \tag{16.6b}$$

图 16.3 显示了以血液为参考阻抗 Z_1 计算得到的各种人体组织的反射系数分贝值。这些系数之间的差别足以分辨不同的组织，这是医学超声成像的基础。

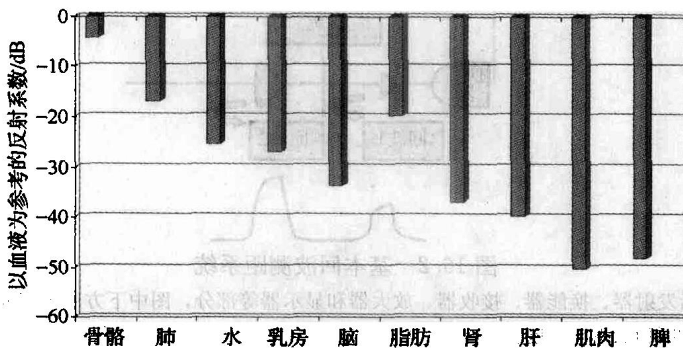


图 16.3 以血液为参考阻抗的人体各种组织的反射系数分贝值

1951 年，英国外科医生 J. J. Wild 和美国电气工程师 J. Reid（现在是一位荣誉退休教授）在明尼苏达用一台废弃的 15MHz 雷达仿真器进行研究，发现了超声对于医学诊断具有的重要价值，当时他们想用超声检测胃癌。他们对脉冲回波的波形进行了各种计算，用于分析健康组织与病变组织的特性，并由此开创了所谓的组织定征（tissue characterization）这个新领域。在此过程中，Wild 和 Reid 开发了近乎实时的超声成像系统，发明了将不同换能器直接放在人体皮肤上的很多种方法。换能器用于发射声波、接收脉冲回波，并将回波信号转换成电信号。

20 世纪 50 年代，国际上很多研究小组开始关注超声用于医学诊断的潜能。日本的 S. Satomura、Y. Nimura 和 T. Yoshida 用多普勒频移技术测量了心脏的血流，而瑞典的 I. Edler 和 C. H. Hertz 则用超声研究心脏的运动，开创了超声心动图（echo-cardiography），研究心脏的动力学特性。

经过世界各国研究人员的不断努力，超声成像技术迅速发展。不过，虽然当时超声已经可以检测人体内部脏器和心脏的活动，但是，其使用仍然只局限于实验室。在 20 世纪 60 年代初之前，X 射线机广泛用于孕妇和胎儿的成像。而英国流行病学家 Alice Stewart 却在 1956 年发表论文指出，儿童患癌症死亡与母亲在怀孕期间受到 X 射线照射有关。这一悲剧事件的发现第一次给超声成像带来了商机，因为它是一种

安全的胎儿成像技术。

20 世纪 60 年代初，好几家公司开发了适用于胎儿等人体内部结构成像的超声成像系统。1965 年，德国西门子公司的 R. Soldner 设计了第一台实时机械超声成像扫描仪。1968 年，I. McDonald 和 T. G. Brown 博士成功地研制了第一台超声诊断仪产品。

这些研究人员最终所偏爱的脉冲回波超声成像仪的基本结构如图 16.4 所示。图像由一系列按照一定顺序排列的脉冲回波扫描线构成，其中每条扫描线对应于图 16.1 所示脉冲序列中的一个脉冲。图 16.4 显示了超声波声束扫描某个椭圆状物体时的情景。首先，将压电换能器置于某个扫描位置，它所发射的声波脉冲遇到椭圆物体前后两个边界时会产生反射波，这两个回波信号在监视器上呈现为相应垂直扫描线上的两个亮点。注意，监视器上显示的扫描线从上至下是时间增加的方向。然后，换能器沿某个方向平移到下一个扫描位置，再重复上述检测过程，依此类推，直到整个物体扫描结束为止。最后，显示在监视器上的这组垂直扫描线就依次对应于换能器的各个扫描位置，从而形成了物体的超声图像。早期这种超声图像要么用长余辉阴极射线管显示，要么用曝光时间较长的照相机拍摄。这种成像模式称为 B 型超声，“B”是“亮度”(brightness) 的意思，因为其回波幅值的大小用灰度显示来表示。B 型超声刚问世时，其监视器只能显示黑白两种颜色，白色对应于大回波幅值，黑色则对应于小回波幅值。后来，随着电子技术的发展，监视器能够用真正的灰度级来表示各种回波幅值的大小，这样，就可以显示出人体组织之间比较细微的差别。

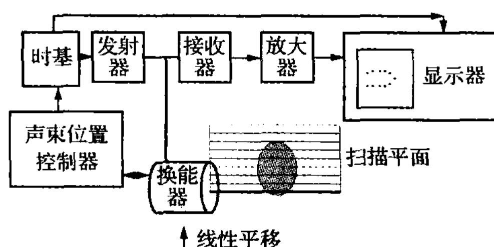


图 16.4 机械扫描式脉冲回波成像系统的基本结构

注：与回波测距系统不同的是，此系统有一个声束位置控制器，并且，信号显示方式也不同。

20 世纪 70 年代，有人开发了电子开关控制的电子激发式微型换能器阵列，于是，机械移动式的单换能器结构立即就被快速实时的成像阵列所取代。J. C. Somer 和 N. Bom 这两位科学家为此作出了重要贡献。Somer 在 1968 年证实了超声相控阵实现的可能性，而 Bom 则于 1971 年在荷兰首次研制成功线性相控阵。到了 20 世纪 80 年代，几乎所有超声成像系统的产品都采用了阵列结构。

从此以后，超声成像经历了许多重要的发展，本章后面的 16.2.6 节和 16.2.7 节会陆续讲述这些新进展。下面先介绍生物组织的声散射和声吸收，以及用于超声成像的压电换能器、换能器阵列、成像系统和各种特殊的信号处理方法，它们是现代超声成像系统的基础。

16.2.2 声波的传播、反射和折射

超声成像基于声波的发射和回波的接收。声波是一种振动波，它可以穿过介质，或者沿介质表面传播。小石块掉进平静的水面时会激起一圈圈扩散的涟漪。同样，

振动波也具有随时间变化的空间传播过程。

平面波和球面波是描述声波传播的两种简单实用的模型。球面波的声波波阵面是一个半径随时间不断扩大的球面。而平面波则是一个一维模型，它假设声波的波阵面在垂直于传播方向的二维平面上是无限大的。简单的一维平面波可以用如下方程表示

$$p(z,t)=A\exp(\mathrm{i}\omega(t-z/c_0))+B\exp(\mathrm{i}\omega(t+z/c_0)) \tag{16.7}$$

式中 t——时间；

z——传播方向上的坐标位置。

式中两个指数函数的参数分别表示沿 z 轴正方向和负方向传播的声波。如果声波的角频率 ω 是常数，那么，当第一个指数函数的参数等于 π 时，该函数的值是个实数，其幅值为-1。然后，随着 t 和 z 的增加，这个指数函数的参数会等于 3π，此时该函数的幅值再一次等于-1。由此可见，指数函数的参数描述的是向前传播的波阵面，波阵面的延时等于距离除以声速 c_0，即 $t=z/c_0$。

波数 k 定义为

$$k=2\pi f/c_0=\omega/c_0 \tag{16.8}$$

式中 f——频率。

在描述声波时，图 16.5 所示的正弦波具有两种含义，它既可以看作某时刻下随传播距离变化的声波振动（其波长为 $\lambda=2\pi/k$），也可以看作某特定位置下随时间变化的声波振动（其周期为 $T=2\pi/\omega$）。

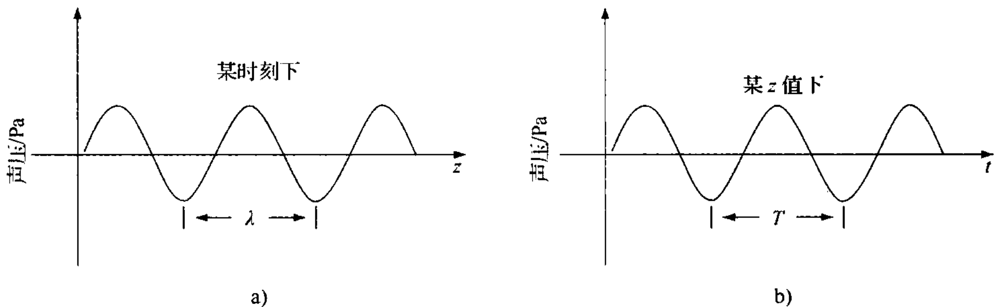


图 16.5 图 a 为某时刻下随空间传播距离变化的声波传播正弦函数，图 b 为某固定距离 z 下随时间变化的声波传播正弦函数

注：图中纵坐标表示声压。两图中两个相邻波谷或者波峰之间的距离或者时间分别定义为声波的波长 λ 和周期 T。

液体中，声压 p 与质点振动速度 v 之比称为声波的比声阻抗（specific acoustic impedance），即特征声阻抗：

$$Z_L=p/v_L=\rho c_L \tag{16.9}$$

式中 ρ——介质的密度。

下标 L 表示纵波，这种波在成像技术中最重要。注意，向前传播的声波 Z_L 值为正，而向后传播的声波 Z_L 值为负。对于 20℃淡水，$\rho=998\mathrm{kg/m^3}$，$c_L=1481\mathrm{m/s}$，

$Z_L=1.48$M Rayl（即 $10^6\,kg/m^2s=10^6$声欧姆）。纵波的振动方向与传播方向一致，其正弦波的正半周称为密波，负半周则称为疏波（如图 16.5b 所示）。

瞬时声强（instantaneous intensity）定义为

$$I_L=pv^*=pp^*/Z_L=vv^*Z_L \tag{16.10}$$

下面利用式（16.7）表示的平面波方程来求反射波的幅度。如图 16.6 所示，假设单频率平面声波在某种理想液体介质中垂直传向两种介质的交界面，声波与交界面的距离为 d。平面波是一种正弦波，它在传播方向 z 之外的 x 和 y 两个方向上视为无限大。此例中，设声波所在介质的波数为 k_1，阻抗为 Z_1，而第二种介质的波数为 k_2，阻抗为 Z_2。如图所示，第二种介质左边的声波可以用向前和向后传播的两个声波的组合来表示，即

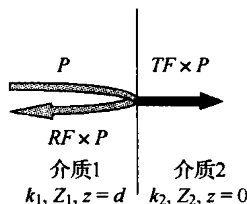


图 16.6　声波在阻抗为 Z_1 和 Z_2 的两种介质交界面处的一维传播模型

$$p=p_0\exp(i(\omega t-k_Lz))+RFp_0\exp(i(\omega t+k_Lz)) \tag{16.11}$$

这是式（16.7）的改进形式，其中 $k_L=k_1$，RF 是负向声波幅值的反射系数。

为了理解反射系数 RF 的求解过程，最好先用人们比较熟悉的电学变量来模拟声波变量，也就是，将声压比作电压，将质点振动速度比作电流。这样，如图 16.6 所示，$z=d$ 处可以看作有一个电压源；第二种介质可以看作位于 $z=0$ 处的阻抗为 Z_2 的一个实负载，它是半无限大。$z=0$ 处的声压 p_2 可以看作 Z_2 两端的电压差。于是，去掉式（16.11）的公共项 exp（$i\omega t$），可得

$$p_2=p_0(1+RF) \tag{16.12}$$

$z=0$ 处的质点速度可以看作流动方向相反的两个电流之和，这两个电流分别对应于入射声波和反射声波。由式（16.9）可知

$$v_2=(1-RF)p_0/Z_1 \tag{16.13}$$

则阻抗 Z_2 就是

$$Z_2=\frac{p_2}{v_2}=\frac{(1+RF)Z_1}{1-RF} \tag{16.14}$$

求解此方程，即可得

$$RF=\frac{Z_2-Z_1}{Z_2+Z_1} \tag{16.15a}$$

透射系数 TF（Transmission Factor）就是

$$TF=1+RF \tag{16.15b}$$

即

$$TF=\frac{2Z_2}{Z_1+Z_2} \tag{16.15c}$$

例 16.2　假设图 16.6 所示的介质 1 是水，请分别计算开放界面、完全匹配界面和刚性界面这 3 种情况下声波的反射系数和透射系数。

解：开放情况下介质 1 与空气交界，这种情况可以用开路电路来模拟，即

$Z_2=0$。由式（16.15a）可知，此时 $RF=-1$，也就是入射声波会产生180°反转，于是，反射声波与入射声波相互抵消，因此 $TF=0$。

阻抗匹配情况下 $Z_1=Z_2$，于是 $RF=0$，也就是没有反射声波；而 $TF=1$，声波全部透射到介质2中。

刚性边界的情况下 $Z_2=\infty$，相当于短路电路，是一个无声压边界，于是 $RF=1$，入射声波将被全部反射回来，且没有相位反转，入射声波与反射声波同相位。此时 $TF=2$。■

液-液界面上的斜入射平面波

如果入射波不与界面垂直，那么，会发生什么情况呢？图16.7描述了这种情况，液体介质1中的单频率纵波在 x-z 平面上与法线成 θ_I 角入射到液体介质1与液体介质2的交界面上。假设界面处的声压和质点振动速度连续，且声波波数的切向分量也相互匹配，即在界面处有

$$k_{1x}=k_1\sin\theta_I=k_2\sin\theta_T=k_1\sin\theta_R \tag{16.16a}$$

式中　k_1 和 k_2——分别是介质1和介质2的波数；

θ_I——入射角；

θ_R——与入射角相等的反射角；

θ_T——折射角。

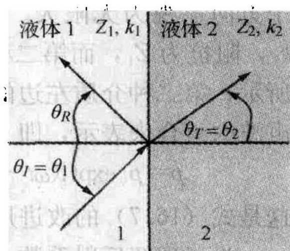


图16.7　液-液交界面处的斜入射纵波（x 轴与界面平行，而 z 轴则与界面垂直）

声学的Snell定理是：

$$\frac{\sin\theta_I}{\sin\theta_T}=\frac{c_1}{c_2} \tag{16.16b}$$

根据此式可以计算折射角 θ_T。式（16.16a）则可用于计算反射角 θ_R。由这些公式可知，z 方向上的各个波数分量分别为

$$\text{入射波}\ k_{Iz}=k_1\cos\theta_I \tag{16.17a}$$

$$\text{反射波}\ k_{Rz}=k_1\cos\theta_R \tag{16.17b}$$

$$\text{透射波}\ k_{Tz}=k_2\cos\theta_T \tag{16.17c}$$

不同入射角和折射角下两种介质的有效阻抗分别为

$$Z_{1\theta}=\frac{\rho_1c_1}{\cos\theta_I}=\frac{Z_1}{\cos\theta_I} \tag{16.18a}$$

和

$$Z_{2\theta}=\frac{\rho_2c_2}{\cos\theta_T}=\frac{Z_2}{\cos\theta_T} \tag{16.18b}$$

此处阻抗是入射角或者折射角的函数。当声波垂直入射时，阻抗最小；随着入射角的增加，阻抗也增大。入射声波进入介质2时，传播方向发生变化，这种现象称为声波的折射。对于两种半无限大液体介质的交界面，各种介质的声阻抗就可以用式（16.18a）和式（16.18b）来表示，于是，根据式（16.15a），声波从介质1到介质2的反射系数就是

$$RF=\frac{Z_{2\theta}-Z_{1\theta}}{Z_{2\theta}+Z_{1\theta}}=\frac{Z_2\cos\theta_I-Z_1\cos\theta_T}{Z_2\cos\theta_I+Z_1\cos\theta_T} \tag{16.19a}$$

声波沿折射角 θ_T 方向传播的透射系数则为

$$TF=\frac{2Z_{2\theta}}{Z_{1\theta}+Z_{2\theta}}=\frac{2Z_2\cos\theta_I}{Z_2\cos\theta_I+Z_1\cos\theta_T} \tag{16.19b}$$

先用式（16.16b）算出 θ_T，然后，就可以求得 RF 和 TF 的值。这种液-液界面常用作组织与组织交界面的声波传播近似模型。

例 16.3 假设声波入射到水与肾组织之间的交界面上，请计算入射角为 45°和 50°时的反射系数。声波在水（即介质 1）和肾组织（即介质 2）中的声速和阻抗分别为 1.48km/s 和 2.05km/s，以及 1.48M Rayl 和 2.89M Rayl（见表 16.1）。

解：

根据 Snell 定理，由式（16.16b）可求得入射角为 45°时的声波折射角为

$$\theta_T=\arcsin[(2.05/1.48)\sin45]=78.4°$$

于是，由式（16.19a）可得入射角为 45°时的反射系数为

$$RF=\frac{2.89\times\cos45-1.48\times\cos78.4}{2.89\times\cos45+1.48\times\cos78.4}=0.746$$

在计算入射角为 50°时的反射系数之前，应该先算出折射波平行于界面传播时的临界入射角。设 Snell 定理中的 θ_T 等于 90°，求得该临界角为 $\theta_I=46.22°$。因此，50°入射角已经超过了临界角，声波会全部反射，没有透射，即 $RF=1$。■

16.2.3 换能器的基本原理

超声系统的关键技术就是声波的产生和接收，这两个功能都由换能器实现。换能器既能够将电信号转换成声波信号，反过来，也能够将声波信号转换成电信号。换能器内部其实是一种压电晶体或者压电陶瓷，外加电压时，它会发生变形而发射出声波；反过来，当晶体受到机械变形作用时，就会产生电荷，因此，换能器可以将返回的声波重新转换成电信号。这种压电效应是居里兄弟于 1880 年发现的。

下面用一个简单模型来描述压电换能器的基本声学特性和电学特性。如图 16.8 所示，设换能器用一片长方形压电晶体制作，其左右两侧装有电极，两个侧面的截面积都为 A，且侧面的上下边要比压电晶体的宽度 d 大很多，至少大 10 倍。由于压电材料是绝缘电介质，因此，它其实构成了一个电容器，电容值为

$$C_0=\varepsilon^S A/d \tag{16.20}$$

式中 ε^S——无变形情况下晶体的介电常数。

晶体是固体，不是液体，其内部会产生弹性波。应力是描述弹性波的一个变量，它是指单位面积上的力，用符号 T 表示。另一个重要的弹性变量是应变，用 S 表示，它等于晶体长度的变化量除以其原始长度，也就是相对长度变化。压电材料在各个不同方向上的应力和应变都不同，不过，对于单向传播的声波，其产生的应力与应变之间具有如下关系

$$T = C^D S - hD \tag{16.21}$$

式中 h——压电常数；

C^D——弹性刚度常数（elastic stiffness constant）。

此方程称为修正的 Hooke 定律，右边第一项是应变引起的应力，第二项则是外加电场引起的应力。弹性刚度常数 C^D是在电介质位移量设定为 D 的情况下测得。如果外加电场为 E，则

$$D = \varepsilon^S E = \frac{\varepsilon^S AV}{dA} = C_0 V/A \tag{16.22}$$

如果在两个电极之间加上一个电压脉冲，那么，换能器两边由压电效应产生的冲激力为

$$F(t) = TA = (hC_0 V/2)[-\delta(t) + \delta(t - d/c)] \tag{16.23}$$

式中 δ——冲激函数，如图 16.8b 所示；

c——声速。

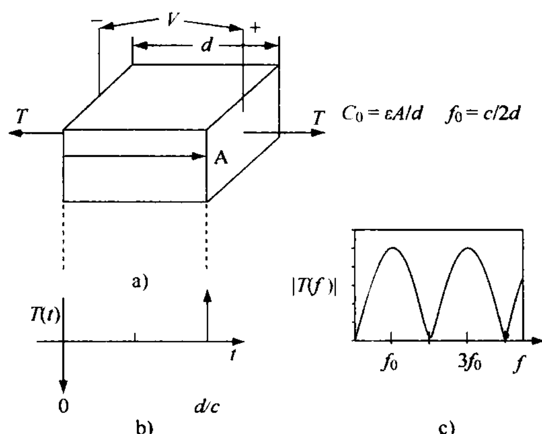


图 16.8 压电晶体向阻抗与之匹配的介质发射声波的示意图（图 a）；应力的时间响应（图 b）；应力的频率响应（图 c）

如果压电晶体两侧的外部介质与晶体换能器具有相同的声阻抗 Z_c；那么，两个电极之间的声速为 $c = \sqrt{C^D/\rho}$。

压电晶体会产生声波，因此，它可以看做一种具有独特音调（即频谱特性）和谐振频率的蜂鸣电容器。将式（16.23）进行傅里叶变换就可以得到此冲激响应的频谱，即

$$F(f) = -i(hC_0 V)e^{\frac{-i\pi fd}{c}} \sin[\pi(2n+1)f/2f_0] \tag{16.24}$$

如图 16.8c 所示，其谐振基波频率为 $f_0 = c/2d$，奇次谐波（即 $n=0$，1，2，3…）处具有最大值。

1. 换能器的电阻抗

换能器两侧产生声压时会在两个电极之间形成一个辐射阻抗（radiation impedance）Z_A，该阻抗与晶体电容的容抗串联，于是，换能器总的电阻抗为

$$Z_T = Z_A - \mathrm{i}(1/\omega C_0) = R_A(f) + \mathrm{i}[X_A(f) - 1/\omega C_0] \tag{16.25}$$

图 16.9a 所示是换能器的等效电路，其中辐射阻抗 Z_A 由实部 R_A 和虚部 X_A 组成。实部 $R_A(f)$ 可以从换能器的总有功功率 W_E 求得。如果在外加电压 V 的作用下流过换能器的电流为 I，则 W_E 为

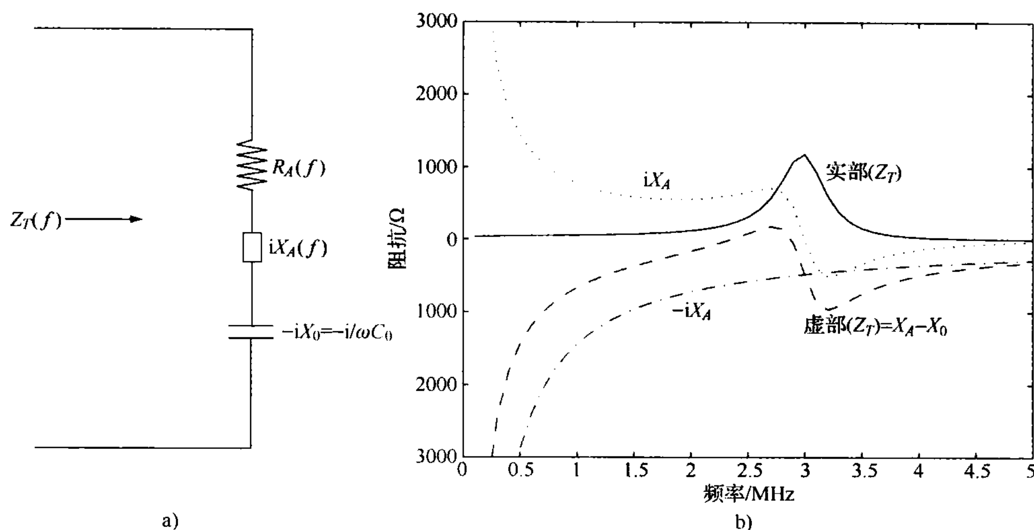


图 16.9 换能器等效电路（图 a）和换能器阻抗随频率变化的函数曲线（图 b）

$$W_E = II^* R_A/2 = |I^2| R_A/2 \tag{16.26a}$$

其中 $I = \mathrm{i}\omega Q = \mathrm{i}\omega C_0 V$，$Q$ 为电荷。假设与晶体阻抗相等的周围介质的修正比声阻抗 (modified specific acoustic impedance) 为 $Z_C = \rho c A$（即包含面积因子的声阻抗），那么，换能器两边所发射的总功率为

$$W_A = ATT^*/(2Z_C/A) = A^2|F(f)/A|^2/2Z_C = |hC_0 V\sin(\pi f/2f_0)|^2/2Z_C \tag{16.26b}$$

令式（16.26a）和式（16.26b）所表示的两个功率相等，就可以求得 R_A 为

$$R_A(f) = R_{AC}\cdot \mathrm{sinc}^2(f/2f_0) \tag{16.27a}$$

其中 $\mathrm{sinc}(x) = \sin(\pi x)/(\pi x)$，且

$$R_{AC} = \frac{k_T^2}{4f_0 C_0} = \frac{d^2 k_T^2}{2A\varepsilon^S} \tag{16.27b}$$

式中 k_T——电声耦合常数（electroacoustic coupling constant)，且 $k_T = h/\sqrt{C^D/\varepsilon^S}$。

注意，R_{AC} 与电容 C_0 和换能器的面积 A 成反比，且直接与厚度 d 的平方成正比。并且，谐振时有

$$R_A(f_0) = \frac{k_T^2}{\pi^2 f_0 C_0} \tag{16.27c}$$

根据网络理论，辐射阻抗的虚部可以通过实部的 Hilbert 变换求得，即

$$X_A(f) = \mathfrak{J}_{Hi}[R_A(f)] = R_{AC}\cdot\frac{[\sin(\pi f/f_0) - \pi f/f_0]}{2(\pi f/2f_0)^2} \tag{16.28}$$

图 16.9b 显示了换能器阻抗随频率变化的函数曲线，谐振频率 f_0处 R_A 达到最大值，而此时 X_A 为 0。

上述简单模型描述了压电换能器的基本特性，在谐振频率处其电阻抗的辐射阻抗实部为最大，声压频谱也在谐振频率处有峰值。

例 16.4 设某块压电陶瓷片 PZT5A 的两面均为边长是 2.5mm 的正方形，其谐振频率为 3MHz，且 $\varepsilon^S/\varepsilon_0=830$，$c=4.35\text{km/s}$，$k_T=0.49$，$\varepsilon_0=8.85\text{pF/m}$。请计算该换能器谐振时的阻抗。

解：首先，计算产生 3MHz 谐振频率的压电陶瓷块所需的厚度。由谐振频率 $f_0=c/2d$，可得 $d=c/2f_0=4.35\text{mm}/\mu\text{s}/2\times3\text{MHz}=0.725\text{mm}$。

然后，计算压电陶瓷的电容

$$C_0=\frac{830\times8.85\times10^{-12}\text{F/m}\times(2.5\times10^{-3}\text{m})^2}{7.25\times10^{-4}\text{m}}=63.3\text{pF}$$

于是，由式（16.27c）可以计算 R_A：

$$R_A=\frac{0.49^2}{\pi^2\times3\times10^6\text{Hz}\times63.3\times10^{-12}\text{F}}=128.2\Omega$$

注意，此式分母的单位是 $\text{Hz}\cdot\text{F}=\text{A/V}=\Omega^{-1}$。

最后，由于谐振时 $X_A=0$，根据式（16.25），可得

$$Z_T=128-\text{i}/(2\times\pi\times3\times10^6\text{Hz}\times63.3\times10^{-12}\text{F})=128-\text{i}838\Omega$$ ■

2. 换能器的频率响应

设计换能器时，首先是要改变其频谱的形状，得到所需的带宽，使换能器具有较窄的冲激响应波形；其次，是要提高换能器的电声效率。电声效率的一种衡量指标是换能器的损耗，也就是，到达人体组织等被测介质的声功率的时间平均值 W_R 与电源输出最大功率 W_g 之比，即

$$TL(f)=\frac{W_R}{W_g} \tag{16.29a}$$

其分贝（dB）形式定义为

$$TL_{\text{dB}}(f)=10\lg TL(f) \tag{16.29b}$$

换能器的损耗还可以进一步分解为电损耗因子 $EL(f)$ 与声损耗因子 $AL(f)$ 之积，即

$$TL(f)=EL(f)AL(f) \tag{16.30}$$

如果要将尽可能多的电功率传输给换能器的辐射阻抗，那么，就要在所需带宽上使 $EL(f)$ 最大化。图 16.10 所示的换能器设计等效电路模型中显示了与这个问题相关的各个参数。图中的换能器电阻抗通过一个可调电感与一个内阻为 R_g 的电压源相连。此时换能器的电损耗为

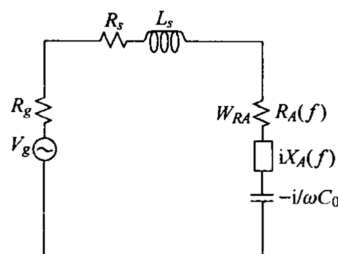


图 16.10 电压源通过可调电感与换能器等效电路相连

$$EL(f)=\frac{4R_A(f)R_g}{(R_A(f)+R_g+R_s)^2+\left(X_A(f)-\frac{1}{\omega C_0}+\omega L_s\right)^2} \tag{16.31a}$$

如果调节可变电感，使电路中的电容与电感串联谐振，即 $L_S=1/(\omega_0^2C_0)$，且设 $R_s=0$，那么，谐振时

$$EL(f_0)=\frac{4R_A(f_0)R_g}{[R_A(f_0)+R_g]^2} \tag{16.31b}$$

注意，如果 $R_A=R_g$，且 $R_s\ll R_g$，那么 $EL(f_0)\approx1$。

声损耗 $AL(f)$ 是频率的一个复杂函数，但谐振时其值的计算比较简单。本节前面所述的换能器简单模型中，为了简化，假设压电晶体两边的声负载具有与压电晶体相同的声阻抗；但是，对于如图 16.11 所示的换能器探头，一般这种情况并不成立。注意，此图中压电晶体的位置与图 16.8 不同，两者相差 90°。图 16.11 的压电晶体上面加了背衬材料，下面则通过透镜与水或者人体组织接触，构成换能器的底部，也就是工作端面。前向声波经工作端面发出，并沿 z 轴正方向传播出去。图中向下的方向就是 z 轴正方向。沿 z 轴负方向传播的反向声波则被背衬材料吸收而抑制，背衬材料还有助于增加换能器的带宽。下面按照图 16.8 从左到右的习惯来叙述。如果设晶体左边的声阻抗为 Z_L，晶体右边沿 z 轴正方向的声阻抗为 Z_R，即 $Z_L=Z_B$（背衬阻抗），$Z_R=Z_W$（水的阻抗）；那么，谐振时其声损耗可以由下式计算

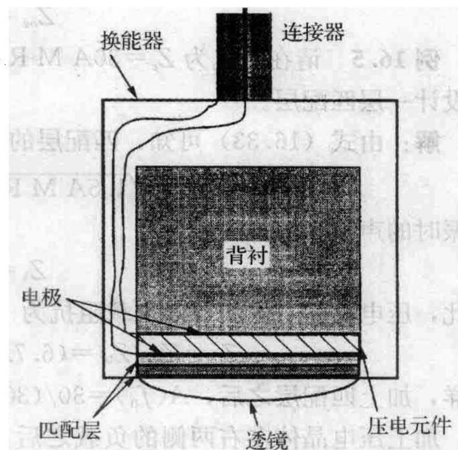


图 16.11 一种单晶体机械式换能器的结构
注：由上至下是前向传播方向。

$$AL(f_0)=Z_R/(Z_L+Z_R)=Z_W/(Z_B+Z_W) \tag{16.32}$$

注意，这些阻抗都包含了面积因子 A。此式的物理意义是：换能器辐射阻抗所接收的电功率被转换为声功率，并从压电材料两侧的电极表面辐射出去时，沿 z 轴正方向传播的前向功率所占的比例就是式中的 $AL(f_0)$。

图 16.11 所示是单晶体的换能器结构，其主要部分有背衬、粘附了导电电极薄层的压电晶体、匹配层以及声阻抗与水接近的透镜（其作用是聚焦，详见 16.2.6 节）。为了扩大换能器的带宽，实际产品的匹配层不止一层，下面还会讨论这个问题。

前述第一个简化模型中，假设压电晶体负载的声阻抗与晶体的声阻抗相等，即 $Z_C=Z_L=Z_R$，因此，就有 $AL(f_0)=0.5$。再举一例，假设图 16.8 晶体左边的负载是

空气，即 $Z_L=0A$ M Rayl⊖；而右边的负载是水，即 $Z_R=1.5A$ M Rayl；那么，此时可达到 $AL(f_0)=1.0$，但是，其代价是换能器的带宽极窄，相应的脉冲响应时间很长。有趣的是，如果背衬材料与晶体相匹配，例如 $Z_L=Z_C=30A$ M Rayl，且 $Z_R=1.5A$ M Rayl（即水），那么，此时 $AL(f_0)=0.048$，几乎不能转换出有用的声波。

根据上述有关声损耗的分析可知，为了提高前向声波的转换效率，通常需要使用多层匹配层。在谐振频率上，匹配层设计为 1/4 波长的厚度，且其阻抗等于需要匹配的两种阻抗的均值，即

$$Z_{ml}=\sqrt{Z_1Z_2} \tag{16.33}$$

例 16.5 请在阻抗为 $Z_c=30A$ M Rayl 的压电晶体与水（$Z_W=1.5A$ M Rayl）之间设计一层匹配层。

解： 由式（16.33）可知，匹配层的阻抗是

$$Z_{ml}=\sqrt{Z_WZ_C}=\sqrt{1.5A\ \mathrm{M\ Rayl}\times30A\mathrm{M\ Rayl}}=6.7A\ \mathrm{M\ Rayl}$$

谐振时的声输入阻抗为

$$Z_1=Z_{ml}^2/Z_2$$

因此，压电晶体右侧的等效负载阻抗为

$$Z_R=Z_{ml}^2/Z_W=(6.7A)^2/1.5A=30A\ \mathrm{M\ Rayl}$$

这样，加上匹配层之后，$A(f_0)=30/(30+30)=0.5$，提高了 10dB 以上。 ■

加上压电晶体左右两侧的负载之后，用如下更通用的换能器辐射阻抗实部在谐振点上的计算公式，就可以将匹配层的作用包括在内：

$$R_A(f_0)=\frac{2k_T^2}{\pi^2f_0C_0}\left(\frac{Z_C}{Z_L+Z_R}\right) \tag{16.34}$$

现在可以将声损耗和电损耗结合起来估计总的换能器损耗。为了便于论述，这里做个简化，假设声损耗在换能器整个带宽上为常数。实际的损耗可以用更完善的等效电路模型精确计算出来。由式（16.30）和式（16.31.a）可得

$$TL(f)\approx\left[\frac{4R_A(f)R_g}{(R_A(f)+R_g+R_s)^2+\left(X_A(f)-\frac{1}{\omega C_0}+\omega L_S\right)^2}\right]\left[\frac{Z_R}{Z_R+Z_L}\right] \tag{16.35}$$

例 16.6 假设电压源内阻为 $R_g=50\Omega$，请计算以下两种情况下例 16.4 所述压电换能器谐振时的损耗：（1）没有调谐电感，$Z_L=Z_B=6A$ M Rayl，且 $Z_R=Z_W=1.5A$ M Rayl；（2）加一个调谐电感和一层匹配层。

解：（1）例 16.4 中 $R_A=128\Omega$，$\frac{1}{\omega C_0}=838\Omega$，因此

$$TL(f_0)=\frac{4\times128\Omega\times50\Omega}{(128\Omega+50\Omega)^2+(838\Omega)^2}\times\frac{1.5\mathrm{A}}{6\mathrm{A}+1.5\mathrm{A}}=6.98\times10^{-3}$$

即 $10\lg(TL)=-21.6\mathrm{dB}$

⊖ 此处 A 表示面积因子，M Rayl 为阻抗单位“兆瑞利”。——译者注

(2) 根据式 (16.34) 和例 16.4 的计算结果 $R_A=128\Omega$。如果加入调谐电感，电感值为 $L_S=1/(\omega_0^2C_0)=838/(2\pi\times3\times10^6)=44.5\mu H$，于是

$$TL(f_0)=\frac{4\times128\Omega\times50\Omega}{(128\Omega+50\Omega)^2+(838\Omega-838\Omega)^2}\times\frac{30A}{30A+1.5A}=0.673$$

即 $$10\lg(TL)=-1.72dB$$ ■

用分贝 (即 dB) 表示带宽是描述换能器频率特性的一种方法，例如，－6dB 带宽是指频谱上幅值等于峰值的－6dB 的两个点之间的频率差，中心频率则用一高一低这两个－6dB 频率来定义，即

$$f_c=(f_{high}-f_{low})/2$$

图 16.12 显示了某个换能器的频率特性及其冲激响应。上图所示为－6dB 带宽及其相关的各个频率数据；下图所示为冲激响应及其包络线，以及包络线的各个宽度数据。

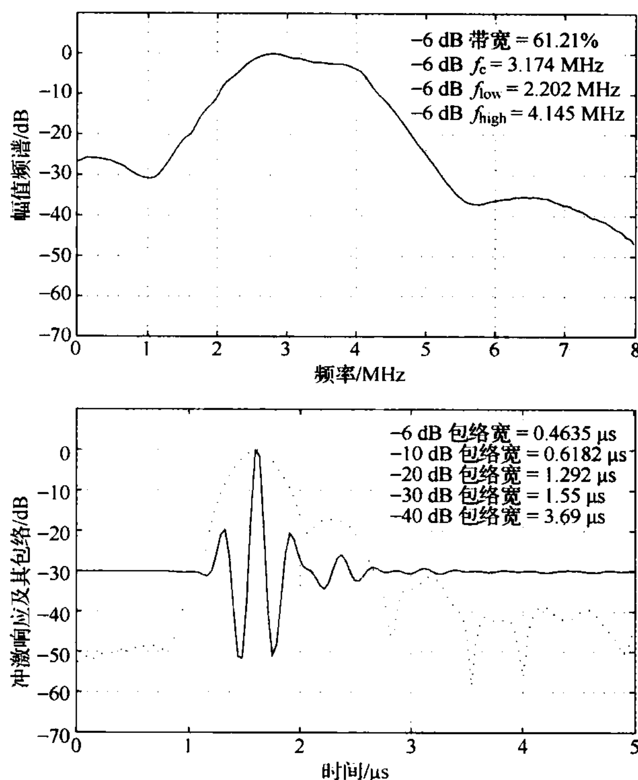


图 16.12 上：换能器的幅值频谱和－6dB 带宽；下：冲激响应及其包络线

3. 换能器的类型

一般超声成像系统并不使用单个晶体换能器，而是采用换能器阵列。例如，一维线阵由直线排列的一组换能器元件组成，其中每个元件可以被单独寻址，或者几

个一组寻址。各个元件均匀排列，元件之间的间距通常设定为水中声波波长的0.5～2倍。图16.13所示就是一种线阵。阵列的主要优点是可以用电子方式实现声束的快速聚焦、定向和切换。而纯粹的机械式单换能器的聚焦是固定的，且换能器只能进行机械平移和转动。不过，就基本工作原理而言，阵列中的每个元件与独立的换能器是一样的。换能器阵列的聚焦问题将在16.2.6节讲述声波衍射时再讨论，下面先介绍阵列的工作原理。

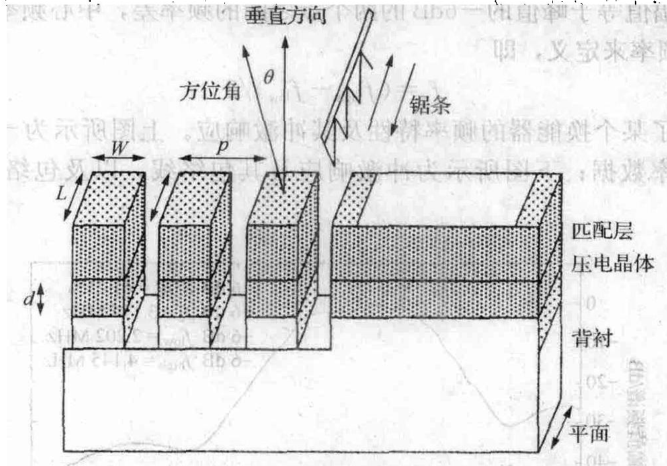


图16.13　用锯子在多层材料上分割出独立的晶体元件，从而制成一维线阵（声波沿垂直方向传播）

线阵和相控阵是两种最常用的超声换能器阵列。如图16.14a所示，线阵通过改变换能器元件的激活窗来扫描产生图像。按照排列顺序，线阵中的元件被一个个依次激活，在每个激活位置上产生一条声波扫描线，并接收其对应于一定测量深度的脉冲回波信号。这样，扫描结束时，共记录到N条平行扫描线的回波信号。相控阵的扫描方式与线阵不同。如图16.14b所示，相控阵激活窗的中心位置固定不变，声波的扫描通过电子定向控制来完成，各条声波扫描线之间相差一个很小的偏转角，形成一组扇形扫描线，每帧图像仍然由N条扫描回波信号构成。图16.15和图16.16分别显示了这两种B型超声阵列所产生的图像。线阵中的换能器元件除了排列成直线形式以外，还可以排列成曲线形式，两种排列形式的工作原理都一样，只是换能器排列在同一平面上还是排列在一个弧形上的区别。图16.17显示了曲线线阵所产生的图像。

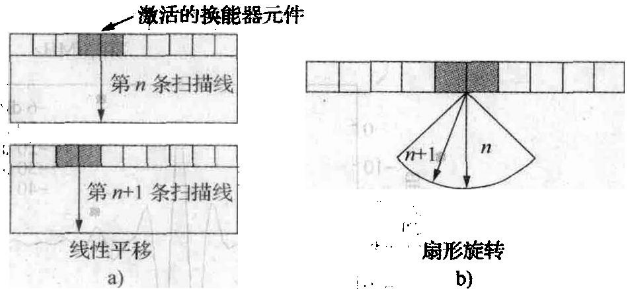


图16.14　换能器阵列的成像原理
a）线阵的平移扫描　b）相控阵的扇形扫描

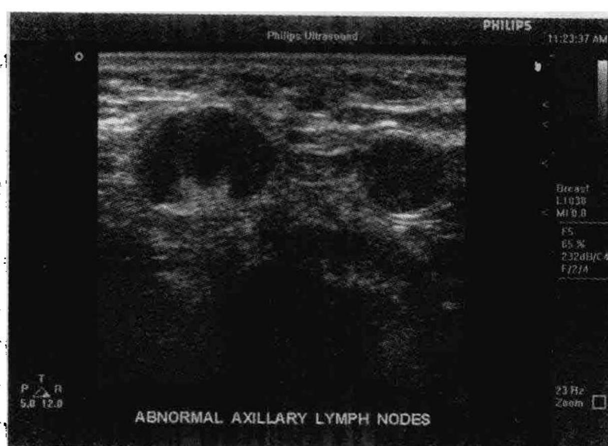


图 16.15 线阵成像示例——12MHz 频率声波的乳腺淋巴结 B 超图像（图片由菲利浦公司赠送）

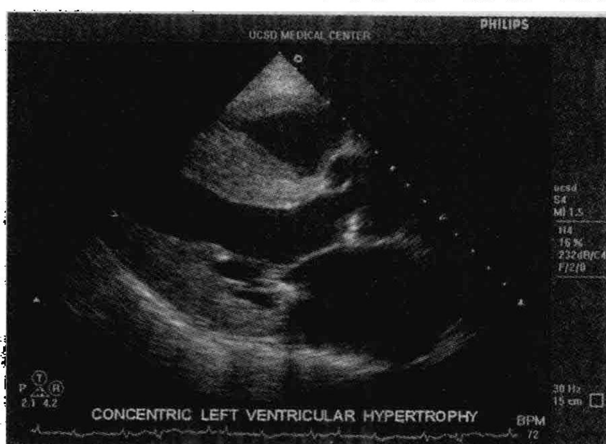


图 16.16 4 相控阵扇形成像示例——4MHz 频率声波的心脏 B 超图像（图片由菲利浦公司赠送）

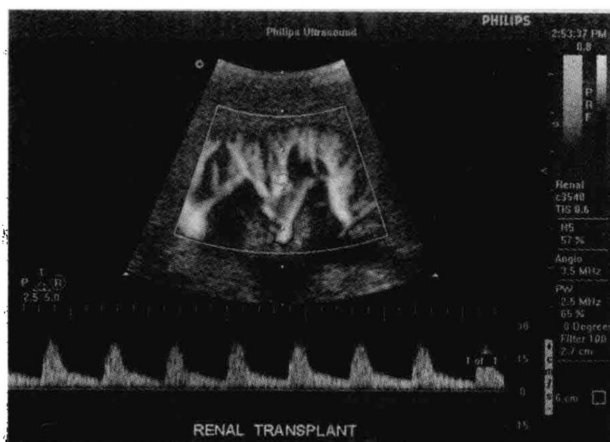


图 16.17 曲面线阵成像示例（上：肾脏的功率多普勒图像；下：相应的脉冲回波多普勒频谱）（图片由菲利浦公司赠送）

例 16.7 对于图 16.14a 所示线阵的平移扫描，假设扫描深度为 $s_d=150\text{mm}$，每帧图像由 100 条扫描线组成，且声波传播速度为 $c_0=1.5\text{mm}/\mu\text{s}$，请计算这种超声图像显示的帧频。

解：每条声波扫描线往返传播所需的时间为：$2\times s_d/c_0=200\mu\text{s}$，因此，包含 100 条扫描线的一帧图像所需的扫描时间共为：100 线/帧×200μs/线＝20ms/帧，也就是图像显示的帧频为 50 帧/s。 ■

如图 16.18 所示，为了满足不同的临床应用需求，超声换能器阵列有各种各样的大小、形状和中心频率。超声检查时，换能器探头一般都通过一层水基凝胶耦合剂与人体表面接触，形成“声波窗”，从体外观察体内的图像。除了声波穿透性很差的骨骼和气体区域以外，通过很小的体表声波窗就能够观察到体内很大的区域。如果这种体表检查方式不能呈现体内某些脏器的清晰图像，那么，还可以使用能够进入食道或者心脏的专用探头，从体内成像。

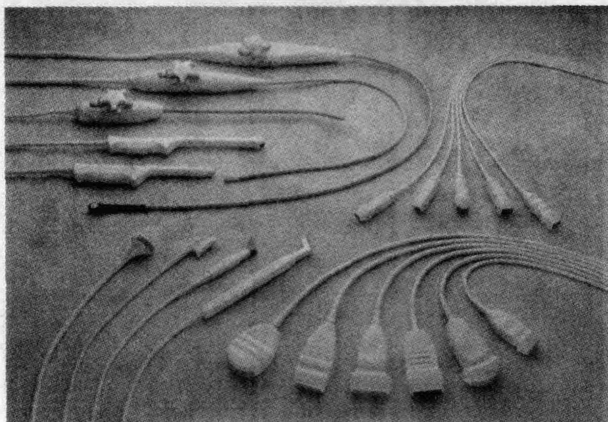


图 16.18 新型超声成像系统配备有适用于各种临床应用的多种换能器探头
注：图中右下方是直线式和曲线式线阵；右上方是相控阵；左边是适用于外科手术的经食道探头和经阴道探头等专用阵列探头（照片由菲利浦公司赠送）。

16.2.4 声波的散射

换能器的声波干涉具有重要的作用。当压电晶体的厚度等于 1/2 声波波长时，晶体会发生谐振，匹配层的厚度一般设计成谐振波长的 1/4。从物体相对于声波波长的大小上可以分析物体对声波的散射作用主要会产生怎样的效应。由于超声成像检测的是人体脏器反射回来的脉冲回波，因此，必须了解物体的大小和形状对于反射回波信号的影响。

如图 16.19 所示，声波的散射包括反射、衍射和漫射 3 种效应。其中，反射前面已经介绍过了。在一种极端情况下，如果物体的尺寸比声波波长大很多，那么，根据 16.2.2 节所述的折射定律，物体表面反射声波的反射角等于入射角。此时，反射

波的幅度由反射系数决定，而反射系数则是入射角的函数。在另一种极端情况下，如果物体的尺寸比波长小得多，那么，此时的反射就是漫射。这时物体与波长相比太小了，以至于其大小和形状对于反射波不产生什么作用。

在这两种极端情况之间，如果物体尺寸与波长相当，那么，就会发生衍射现象。假设将物体表面分成许多无限小的点，并且将这些点与观察点之间用直线连起来，那么，这些直线路径长度之间的差别在声波波长数量级上是很明显的，它们会引起沿这些路径传播的声波之间的明显相位差，并使声波在观察点处产生增强或抵消的干涉效应。此干涉效应形成的复杂的声波散射干涉图案与物体的形状、取向以及物体与观察点之间的距离都有关。即使对于球体或柱体之类简单形状的物体，如果其形状和尺寸发生了与波长数量级相当的变化，那么，在观察点处所产生的散射干涉图案也会随之发生变化。

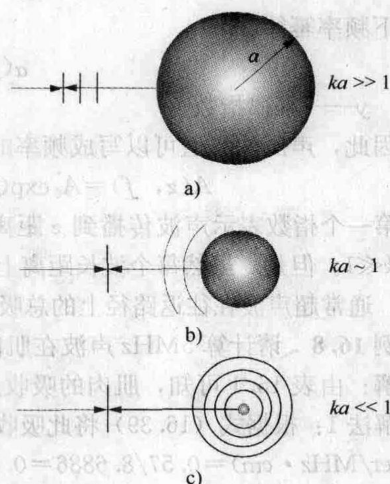


图 16.19 声波的散射

a）比波长大得多的物体的反射 b）与波长大小相当的物体的衍射 c）比波长小得多的物体的漫射

人眼的视觉图像主要依赖于物体的反射光，对应于物体尺寸比光的波长大得多的情况。超声波图像根据观察点到物体各个部分的声波传播直线上的脉冲回波信号成像。与光学成像不同，物体的超声波图像与观察方向和声波的频率都有关。

16.2.5 声波的吸收

声波在实际介质中传播时，会产生损耗。就像作用力中的摩擦力、压力和应力一样，声波在传播时也会向介质传输能量，并产生微弱的局部加热效应。这种损耗称为衰减，可以用距离的指数函数来描述。对于单频率 f_c 的平面波，就是将其幅值乘以一个损耗因子，即

$$A(z,\ t)=A_0\exp(\mathrm{i}(\omega_c t-kz))\exp(-\alpha z) \tag{16.36}$$

式中 α——衰减系数（即吸收系数），常用单位为 neper/cm，即奈培/cm。

另一个常用的幅度测量单位是分贝（dB），是指幅度 A 与 A_0 之比的对数值，即

$$\text{比值(dB)}=20\lg(A/A_0) \tag{16.37}$$

如果声强 I 与幅度 A 的平方直接成正比，即 $I_0\propto A_0^2$，则

$$\text{比值(dB)}=10\lg(I/I_0)=10\lg(A/A_0)^2 \tag{16.38}$$

于是，上述衰减系数α常用 dB/cm 的形式表示，即

$$\alpha_{\mathrm{dB}}=1/z\{20\times\lg[\exp(-\alpha_{\mathrm{neper}}z)]\}=8.6886(\alpha_{\mathrm{neper}}) \tag{16.39}$$

实际测量数据表明，声波的吸收损耗是频率的函数（参见表 16.1）。这类衰减通常遵

循以下频率幂律：

$$\alpha(f)=\alpha_0|f|^y \tag{16.40}$$

式中　y——幂律指数。

因此，声压的幅值可以写成频率的函数，即

$$A(z, f)=A_0\exp(-i2\pi fz/c)\exp[-\alpha_0|f|^y z] \tag{16.41}$$

式中第一个指数表示声波传播到 z 距离时的延时。每个声波波数上的实际损耗极小，即 $\alpha/k\ll1$。但是，虽然每个波长距离上的损耗非常小，很多波长上积累的吸收效应却很强。通常超声波在往返路径上的总吸收量决定了它能够成像的人体组织的深度。

例 16.8　请计算 5MHz 声波在肌肉中传播 10cm 所引起的吸收损耗。

解：由表 16.1 可知，肌肉的吸收系数 $\alpha_0=0.57$dB/cm・MHz。

解法 1：根据式（16.39）将此吸收系数转换成以奈培（neper）为单位的数值，即 α(neper/MHz・cm)＝0.57/8.6886＝0.0656。将该值代入式（16.41）可得 $A/A_0=\exp$（$-$0.0656 neper/MHz・cm×5MHz×10cm)＝3.76×10^{-2}。再转化为分贝形式，即为 $20\lg(A/A_0)=-28.5$dB。

解法 2：如果只需要对数形式的 dB 值，那么，可以直接计算吸收系数，即 $\alpha_0 fz=$ $-$0.57dB/MHz・cm×5MHz×10cm＝$-$28.5dB。■

16.2.6　声波的衍射

实际上，压电换能器所发射的声波并不是平面波，而是具有复杂形状的声波。如图 16.20 所示，从换能器表面小孔不同位置发出的声波及其相互之间的干涉形成了这种复杂声波，这种现象称为衍射。当发射孔径的大小与声波波长数量级相当时，就会产生衍射。

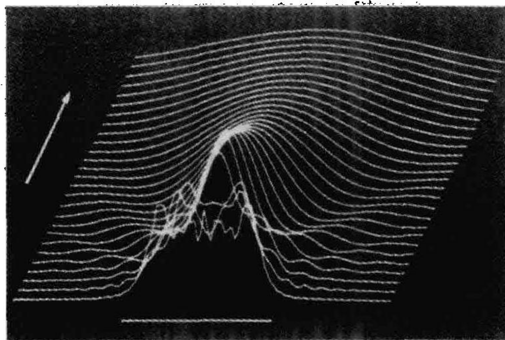


图 16.20　沿 x 轴方向的 40 个波长宽的矩形发射孔所产生的声波衍射场

注：垂直坐标轴表示声强。z 轴的尺度被压缩，其显示范围为 1920 个波长，每条声强分布曲线之间的间隔为 70 个波长。

图 16.21 所示的圆形和矩形小孔是最常见的换能器发射孔的形状。三维空间上的声束分布通常用各个切平面上的声强曲线来描述。对于对称的圆形发射孔，任何垂直于声束中心轴（即 z 轴）的平面上声强分布的形状都一样。而对于矩形发射孔，垂直于声束中心轴的各个平面上的声强分布形状都不同，最重要的成像平面是 x-z 平

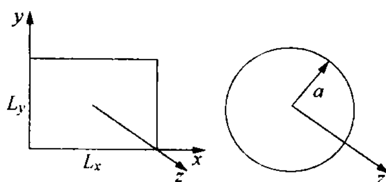


图 16.21　左：矩形发射孔，其平行于 x 轴和 y 轴的边长分别为 L_x 和 L_y；
右：半径为 a 的圆形发射孔（图中 z 轴垂直于小孔所处的 x-y 平面）

面和 y-z 平面。图 16.20 所示是矩形孔 x-z 平面上的声强分布。

不同形状换能器发射小孔发出的声束都具有明显的特征。借用地图等高线的绘制法，可以绘出声压的等高线，声压幅值通常用各个声波扫描深度下的最大幅值归一化之后的 dB 值来表示。其中 −6dB 的等高线很重要。垂直于声束轴的横截面上的声压分布称为声束图。声束图上两个 −6dB 声压点之间的宽度称为半高宽（Full-Width-Half-Maximum，FWHM）。有趣的是，换能器发射孔所发出的声束在某个区域会变细，也就是半高宽变小。沿着声束传播的轴向，靠近发射小孔的区域声压呈现振荡现象，最后一个声压波峰出现的深度称为转变深度，也就是自然聚焦长度 F_N。以转变深度为界，可以将声束场分成两个区域。一个是靠近发射小孔的区域，称为近场，该区域的声束含有振荡的波峰和波谷；另一个是远离发射小孔的区域，称为远场，该区域的声压幅值随着距离的增加逐渐减小，而声束的宽度则逐渐增大（见图 16.20），也就是声束逐渐发散。对于半径为 a 的圆形发射小孔，其转变深度是

$$z_t = a^2/\lambda \tag{16.42a}$$

对于 x 轴方向长度为 L_x 的矩形发射小孔，其在 x-z 平面上的转变深度为

$$z_t \approx L_x^2/(\pi\lambda) \tag{16.42b}$$

自然聚焦长度是发射小孔到声束轴方向上出现的最后一个声压波峰之间的距离，它与转变深度一致。

矩形孔远场声束的形状是小孔截面上声压幅值的傅里叶变换。在均匀发射的情况下，即为

$$A(x_0,\ y_0,\ 0) = \Pi(x_0/L_x)\Pi(y_0/L_y) \tag{16.43a}$$

式中

$$\Pi(x/L) = \begin{cases} 0 & |x| > L/2 \\ 1/2 & |x| = L/2 \\ 1 & |x| < L/2 \end{cases} \tag{16.43b}$$

x-z 平面上的远场声束形状是一个 sinc 函数

$$p(x,z,\omega)\ \frac{L_x\ \sqrt{p_0}}{\sqrt{\lambda z}}\mathrm{e}^{\mathrm{i}\pi/4}\ \frac{\sin(\pi L_x x/\lambda z)}{(\pi L_x x/\lambda z)} = \frac{L_x\ \sqrt{p_0}}{\sqrt{\lambda z}}\mathrm{e}^{\mathrm{i}\pi/4}\,\mathrm{sinc}\left(\frac{L_x x}{\lambda z}\right) \tag{16.44a}$$

图 16.22 的实线显示了这种声束形态的曲线。

对于发射均匀声压 p_0 的圆形小孔，其远场声束的形状是圆形对称小孔函数的二

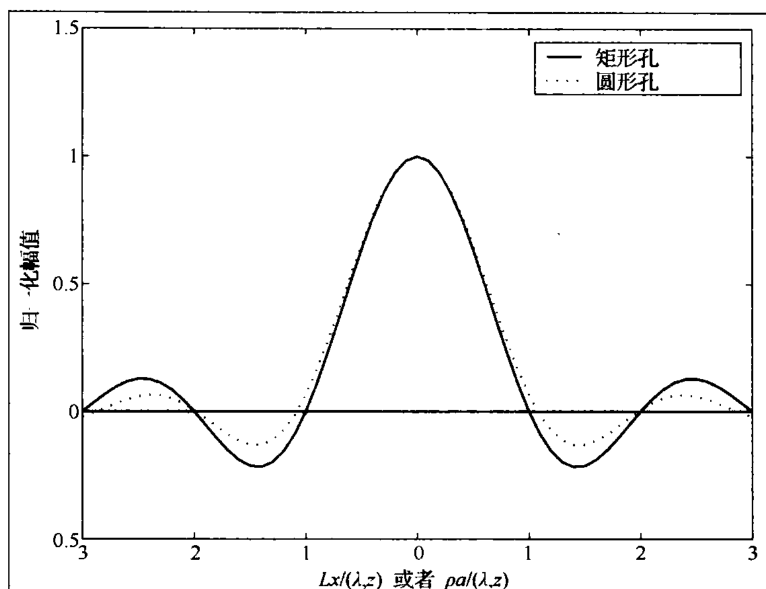


图 16.22 实线是矩形孔远场声束形状的 sinc 函数，虚线是圆形孔远场声束形状的 jinc 函数

注：图中数值是按照矩形孔声束声压的最大值进行归一化后的幅值，两个孔的面积相等。

维傅里叶变换，即

$$p(\bar{\rho},z,\lambda)\approx\frac{\mathrm{i}p_0\pi a^2}{\lambda z}\frac{2J_1(2\pi\bar{\rho}a/(\lambda z))}{2\pi\bar{\rho}a/(\lambda z)}=\mathrm{i}p_0\left(\frac{\pi a^2}{\lambda z}\right)jinc\left(\frac{\bar{\rho}a}{\lambda z}\right)\tag{16.44b}$$

式中 J_1——第一类 Bessel 函数，即 $jinc(x)=2J_1(2\pi x)/(2\pi x)$；

$\bar{\rho}$——观察点（$\bar{\rho}$，z）的径向距离。

图 16.22 的虚线显示了圆形孔远场声束的形状。注意，远场声束的形状不随距离而变，但其幅值随距离下降，且声束的宽度随距离增加。

从这两种远场声束形状上很容易计算声束的 FWHM 宽度，矩形孔 x-z 平面和 y-z 平面上的 FWHM 为

$$\mathrm{FWHM}=1.206\lambda z/L\tag{16.45a}$$

式中 L——相应平面上的孔径。

同理，圆形孔的 FWHM 为

$$\mathrm{FWHM}=0.7047\lambda z/a\tag{16.45b}$$

例 16.9 为了理解图 16.22 中的归一化数据，请计算与矩形孔面积相等的圆形孔的 FWHM 值，并将其与矩形孔的 FWHM 值进行比较。

解： 令 $\pi a^2=L^2$，即 $a(m)=L(m)/\sqrt{\pi}$，代入式（16.45b），可得

$$\mathrm{FWHM}=0.7047\lambda(m)z(m)(\sqrt{\pi}/L(m))=1.249\lambda(m)z(m)/L(m)$$

该值很接近式（16.45a）求得的矩形孔的值 $\mathrm{FWHM}=1.206\,\lambda z/L(m)$。 ■

使用几何聚焦方法可以调节声束变细的位置，使声束在不同探测深度处聚焦。

与光束的聚焦一样，声束聚焦也可以用透镜来实现。不过，与光学聚焦不同的是，用于声波聚焦的透镜既可以做成凹透镜，也可以做成凸透镜，这取决于制作透镜的材料的声速比水（即生物组织）的声速大还是小，并且这两类材料都很容易得到。由几何光学原理可知，光线会在焦点处聚焦。换能器晶体表面贴上声透镜之后，根据透镜组成像的倒数定律，超声探头的总聚焦长度变为自然聚焦长度与透镜几何聚焦长度两者的结合，即

$$1/F_{总}=1/F_N+1/F \tag{16.46}$$

此式表明总聚焦长度比原有的自然聚焦长度要短，也就是，声轴上声压波峰的位置离发射孔更近。例如，假设自然聚焦长度为100mm，透镜的几何聚焦长度为50mm，那么，总的等效聚焦长度就是33.3mm。贴上透镜之后，聚焦平面上声束的形状与不加透镜时的远场声束形状相同，因此，式（16.44）～式（16.45）照样可以用，只是式中的 z 变成了 F。

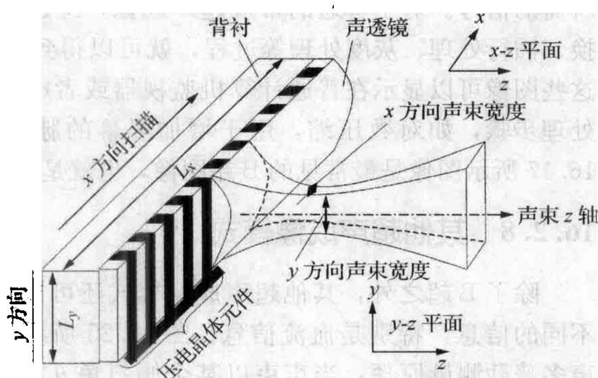


图16.23 一维相控阵在 x-z 平面和 y-z 平面上的聚焦（摘自 Panda，1998）

以上介绍的是固定式压电晶体换能器声波发射小孔的特性。换能器阵列的发射孔可以看作移动式的空间采样小孔。其实，阵列中单个采样小孔所发射的声束与固定式小孔是一样的。如图16.23所示，一维线性阵列实际上有两种聚焦，一是水平面（即 x-z 平面）上的电子聚焦，二是垂直平面（即 y-z 平面）上的固定透镜产生的机械聚焦。

16.2.7 超声成像系统

通过学习成像系统的工作原理可以了解超声图像的形成过程。16.2.1节已经介绍过由单个换能器组成的成像系统，那种系统中的换能器通过机械式或者电子式扫描，依次采集一组声束扫描回波信号线，然后用于成像。这里要介绍的则是图16.24所示的由计算机（即中央处理器（CPU））控制的换能器

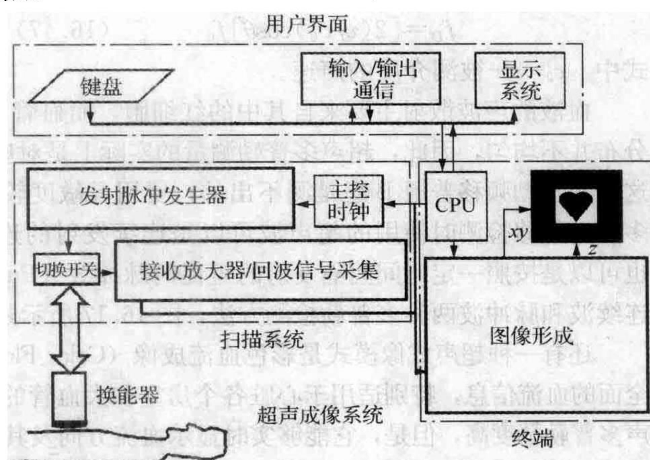


图16.24 数字超声成像系统的框图

阵列成像系统。设置好扫描深度和扫描模式之后，这种系统就在主控时钟的同步控制下，在每个扫描时间间隔上发射一个电脉冲信号。该脉冲信号再触发"发射脉冲发生器"产生一组发射脉冲，然后由这组脉冲去激发阵列中的某一小组换能器元件。其中各个发射脉冲之间具有一定的相对延时，从而形成电子透镜的聚焦作用，使元件组发出的声束产生聚焦，并且可以调节声束在水平面上的发射方向。

换能器发射的声波穿过人体组织时，不断地从各个组织的分界面上产生非均匀散射，由此形成的脉冲回波信号传回到换能器阵列，并被换能器接收。为了减少人体组织衰减作用对成像的影响，用户可以根据成像需要，调节随深度变化的时间增益补偿放大器（Time Gain Compensation，TGC），使不同深度的回波幅值变得均匀一致。回波信号采集器利用电子动态聚焦和滤波等信号处理生成特定超声成像模式所需的信号。对于普通的常规超声成像，经过脉冲回波信号包络线的提取、扫描变换、插行处理、灰度处理等过程，就可以得到图 16.15～图 16.17 所示的灰度图像，这些图像可以显示在普通计算机监视器或者电视机屏幕上。另外，还可以添加其他处理步骤，如对数压缩，用于增加屏幕的脉冲回波幅值显示范围。图 16.15～图 16.17 所示图像是最常见的 B 超图像，也就是亮度模式的超声成像。

16.2.8 其他超声成像模式

除了 B 超之外，其他超声成像模式还可以提供不同的信息，特别是血流信息。图 16.25 所示是超声多普勒测量原理，当声束以某个倾斜角 θ 穿过血流速度为 v 的血管时，超声波频率会产生明显的变化。

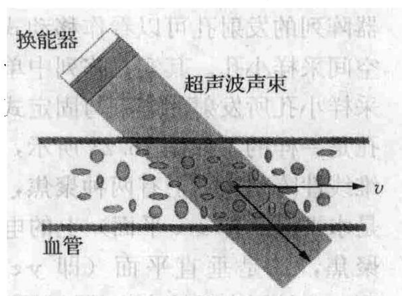


图 16.25 声束以 θ 倾斜角穿过血流速度为 v 的血管

假设声波的原频率为 f_0，那么，透射声波的频移（即所谓的多普勒频移）f_D 就是

$$f_D=[2(v/c_0)\cos\theta]f_0 \qquad (16.47)$$

式中 c_0——被测介质的声速。

血液的声波散射主要来自其中的红细胞，而血管中的血流速度在整个横截面上的分布并不均匀，因此，超声多普勒测量的实际上是对应于某个声速范围的多普勒频谱。这种多普勒频移普通 B 超是测不出的，要用灵敏度较高的超声多普勒仪器才能测量。多普勒频移检测时使用的超声波可以是连续发射的连续波（Continuous Wave，CW），也可以是按照一定时间间隔发射的短促的脉冲波（Pulsed Wave，PW），于是就产生了连续波和脉冲波两种多普勒检测方法，图 16.17 所示是脉冲波多普勒图像。

还有一种超声成像模式是彩色血流成像（Color Flow Imaging，CFI），它可以获得更全面的血流信息，特别适用于心脏各个房室和大血管的成像。这种成像模式虽然没有超声多普勒精度高，但是，它能够实时显示血流方向及其大概的速度。血流方向是指流向换能器还是离开换能器。其图像通过色彩映射关系来表示血流速度的大小和方向。在这种成像模式下，超声探头沿同一个方向发射声束扫描线，并采集血流随时间变化的信号，

然后经过信号处理提取血流速度信息，用于显示。另有一种功率多普勒成像，如图 16.17 上方所示，它只包含血流速度的幅度信息，并不包含血流的方向信息。

一般超声成像都是二维的，显示的是扫描平面上所建立的图像。如果扫描的是一个容积，而不仅仅扫描单个平面，那么，就可以获得三维（3D）图像。其中一种实现方法是，使一维换能器阵列沿着垂直于成像平面的方向进行机械式平移，并在每个成像平面停留足够长的时间，采集每幅平面图像，然后，在各个平面图像之间进行插值计算，就可以获得整个扫描容积的成像数据。3D 图像有多种显示方法，可以显示选定的各个断面，也可以显示物体表面，或者显示体绘制图。体绘制可以通过器官与组织之间的透明分界面显示出器官内部的细节。例如，超声 3D 图像的一种常规应用是利用胎儿皮肤与母亲羊水之间的分界面来显示胎儿的外表，图 16.26 所示的就是这种表面图像。除了一维换能器阵列的机械式平移扫描之外，通过其他机械式扫描方法，或者使用二维静态换能器阵列的全电子式扫描，也可以获得三维容积的超声成像数据。



图 16.26 胎儿头部外表的三维超声成像（照片由菲利浦公司赠送）

16.3 磁共振成像

16.3.1 概述

借助于人体的高含水量，磁共振已经被成功应用于人体的医学成像。人体中水和脂肪所包含的氢原子总量约占体重的 60%。每个氢原子核都有一个带正电荷的质子，因此，氢原子核自旋时会在其周围产生一个很小的磁场并具有磁矩。当氢原子处于强大的静磁场中时，其原子自旋就好像一个具有拉莫尔（Larmor）进动频率的微型陀螺的旋转，拉莫尔频率是物体的一种特性。磁共振成像检测的是物体在垂直于外加静磁场的平面上所产生的一种射频交变磁场，该磁场的交变频率就是物体的

拉莫尔频率。磁共振检测时，物体的原子核首先受到外加磁场的激发，随后，当外磁场撤除时，受激原子核的磁矩就会衰减到平衡状态，并发出电磁波信号。此时用产生外加磁场的同一个线圈检测该信号的电压，就可以测得两个弛豫时间常数（relaxation constant）。其中，一个是纵向磁化弛豫时间常数 T_1，它对组织的热特性较敏感；另一个则是横向磁化弛豫时间常数 T_2，它可以反映不同区域磁场的非均匀性。这两个时间常数都可以用于区分不同的人体组织并用于成像，而 T_1 加权像则比较常用。

如今，磁共振成像（MRI）已广泛用于各种疾病的检测和手术方案的制定，其图像可以清晰地显示出人体内部的解剖细节。这种图像被称为参数化图像，因为需要凭借高技术调整仪器，才能获得有用的图像，突显出各种不同组织之间的对比度差别，用于区分各种不同脏器，识别健康组织和病态组织。MRI 几乎适用于人体全身所有部位，包括颅脑、腹腔、心脏、大血管、乳房、骨骼、软组织、关节、软骨、肌肉以及头部和颈部，用于检查癌症、肿瘤和出血等疾病。并且成人和儿童都可以用。

MRI 的基础是核磁共振（Nuclear Magnetic Resonance，NMR），最早观察到 NMR 现象的是 Felix Bloch 和 Edward Purcell 等人。他们发现，不仅进动的原子核会发出射频（Radio Frequency，RF）信号，而且，利用外加射频磁场还可以控制原子核产生具有拉莫尔共振频率的进动，一旦外加磁场消失，原子核就会发射具有该频率的可测 RF 信号。他们两人的工作成果获得了 1951 年诺贝尔奖。

从此以后，人们开始利用不同化合物所具有的特定共振频率位移来测定物质的组成，并最终进行了生化物质的 NMR 实验，不久就获得了磷、碳和氢等原子核的详细核磁共振频谱图。而且还为人体各个部位的 NMR 研究设计了专用磁体。

Paul Lauterbure 最早意识到 NMR 原理可用于成像。1973 年，他发表了一幅非均匀物体的 NMR 图像。通过改变磁场梯度，他利用 NMR 的 RF 信号测定了物体的空间结构。20 世纪 70 年代中后期，又有人成功地获得了动物和人体的早期 MRI 图像。最初，由于信号太弱，这种成像技术只限于实验室使用。1971 年，Raymond Damadian 证实，恶性肿瘤与正常组织的弛豫常数 T_1 和 T_2 是不同的。另外，Peter Mansfield 还建立了数学模型，用于分析外加强磁场作用下人体内部磁共振信号的变化，并且开发了一种快速成像方法。在这些研究工作的不断推进下，有关 MRI 临床诊断成像的重要基础知识已经形成，现代新型 MRI 仪器已经具有很高的信噪比。在此过程中，Lauterbur 和 Mansfield 对 MRI 作出了重要的贡献，两人共同分享了 2003 年的诺贝尔医学奖。

16.3.2 磁场与电荷

为了解释 MRI 的工作原理，下面先回顾一下磁场的几个重要性质，特别是电荷与磁场之间的相互作用关系。爱因斯坦曾经指出，将电场和磁场看作能量的两个方面是很有用的。

下面分 4 种情况逐步介绍有关 MRI 原理的某些知识。第一种情况：当交流电经过导线时，其周围就会产生磁场。如图 16.27 所示，对于无限长导线，由 Biot-Savart 定律可知，导线中的电流会产生环绕导线的环形磁场，且其磁通量为

$$B_{\phi}=\frac{\mu_0 I}{2\pi r} \tag{16.48}$$

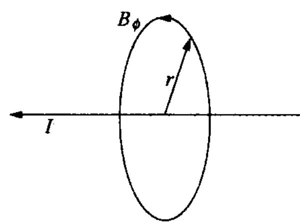


图 16.27 无限长导线中流过的电流所产生的环形磁场

式中 B——磁通量，单位为韦伯（即 Wb），$1\text{T}=10^4\text{Wb/m}^2$，T 为特斯拉；

μ_0——真空磁导率，$\mu_0=4\pi10^{-7}\text{H/m}$（即 Wb/A·m），H 为亨利；

I——电流，单位为 A；

r——离开导线的径向距离，单位为 m。

根据右手螺旋法则，拇指表示电流方向，则四指就表示磁场 B 的方向。

例 16.10 假设导线中通过的电流为 20A，请计算距离导线 5mm 处的磁场强度（即磁通量），单位为 T。

解： 由式（16.48）可得

$$B_{\phi}=\frac{4\pi10^{-7}\times20}{2\pi\times5\times10^{-3}}(\text{Wb/m}^2)\times1\text{T}/10^4(\text{Wb/m}^2)=8\times10^{-6}\text{T}$$ ■

如果将导线绕成一个圆圈，那么，就会产生具有南北两极的磁偶极子。如图 16.28 所示，这是要讲述的第二种情况，一个旋转的电荷就会产生这种情况。设线圈半径为 r，其中流过的电流为 I，取长为 dl 的小段导线，并设这段导线的等效电荷为 q，质量为 m，该等效电荷沿导线圆轨道运行的频率为 ν；那么，电荷产生的磁偶极矩为等效电流与圆圈面积的乘积，即

$$\hat{\boldsymbol{\mu}}=(q\nu A)\hat{z}=q\nu\pi r^2\hat{z}=\frac{1}{2}qr^2\omega\hat{z} \tag{16.49}$$

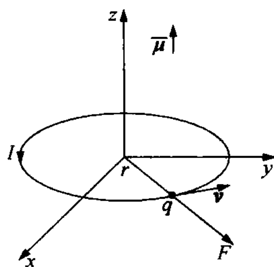


图 16.28 圆形轨道中运行的电荷所产生的磁偶极矩

$\hat{\boldsymbol{\mu}}$ 的方向根据右手螺旋法则确定，就是其单位矢量 $\hat{z}$ 的方向。矢量既有大小又有方向。此处，单位矢量 $\hat{z}$ 的幅值为 1，方向是 z 轴。如果电荷的质量为 m，则其轨道角动量是

$$L=mr^2\omega \tag{16.50a}$$

旋磁比（gyromagnetic ratio）定义为

$$\gamma_c=\frac{\mu}{L}=\frac{q}{2m} \tag{16.50b}$$

由此可见，绕轨道运行的电荷所产生的磁场效应与具有南极和北极的小磁体一样。

那么，反过来，运动的磁场是否可以在导线中产生电流或电压呢？这就是第 3 种情况。如图 16.29 所示的实验装置中，半径为 d 的线圈垂直于 x 轴，强度为 B_0 的永磁

体以恒定的角频率 ω 绕 z 轴旋转；于是，磁体轴线与 x 轴之间的夹角为 $\theta=\omega t$，旋转磁场相对于 x 轴的强度为 $B_0\cos\omega t$，线圈的面积为 $S=\pi d^2$，面积矢量的方向垂直于线圈平面。那么，根据法拉第定理，旋转磁体在线圈中产生的电压 V 与线圈所包围面积中磁通量对时间的变化率成正比，即

$$V=-\frac{\mathrm{d}}{\mathrm{d}t}(B_0\pi d^2\cos\omega t)=-\omega B_0\pi d^2\sin\omega t \tag{16.51}$$

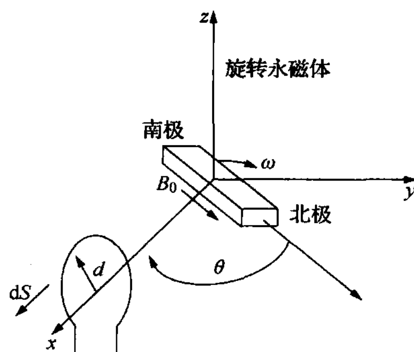


图 16.29 强度为 B_0 的旋转永磁体会在半径为 d 的线圈中产生感应电压

可见，旋转磁体产生的电压 V 是一个正弦交变电压，当磁体轴向与线圈平面方向垂直时，电压达到最大值；当两个方向平行时，电压为零。

第四种情况如图 16.30 所示，旋转的电荷被置于强静磁场 B_0 中。此时，磁场在电荷上产生的作用力可以用如下洛伦兹（Lorentz）方程来表示

$$\hat{\boldsymbol{F}}=q\hat{\boldsymbol{v}}\times\hat{\boldsymbol{B}},\ 即|F|=q|v||B|\sin\theta \tag{16.52a}$$

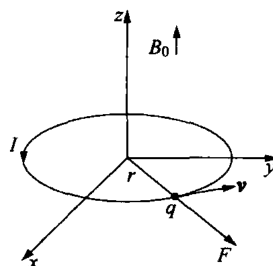


图 16.30 在静磁场中旋转的电荷所受到的洛伦兹力

注：图中标出了电荷的速度矢量 v 和作用力 F，以及磁场 B_0。

其中，速度矢量 $\hat{v}$ 与电荷运行轨道相切，作用力的方向与 $\hat{v}$ 和外加磁场 $\hat{\boldsymbol{B}}_0$ 都垂直，$\hat{v}$ 与 $\hat{\boldsymbol{B}}_0$ 之间的角度为 $\theta=90°$。此时作用力的幅值可以写为

$$|F|=q|v||B|\sin\theta=mv^2/r \tag{16.52b}$$

电荷的角频率 ω_c 定义为电荷旋转速度与半径之比，由上式可知，为

$$\omega_c=v/r=v/(mv/qB_0)=(q/m)B_0 \tag{16.53a}$$

根据式（16.50b）的旋磁比定义可以推导出 MRI 中常用的重要频率公式：

$$\nu_c=\frac{\omega}{2\pi}=\frac{(2\gamma_c)B_0}{2\pi}=\frac{\gamma_c B_0}{\pi} \tag{16.53b}$$

此式表明，电荷的轨道频率 ν_c 正比于外加磁场。不幸的是，这并不是 MRI 真正需要的理论基础，因为上述经典电磁场理论所讨论的电荷不具有绕自身轴的旋转运动。而 MRI 中感兴趣的电荷是电子，它具有自旋运动，就像地球在绕太阳公转的同时具有自转一样。为了获得描述这种情况的重要方程，就需要了解有关自旋状态的量子力学理论。

16.3.3 原子核的自旋状态

由上述理论可知，电子绕自身轴的旋转也会产生微小的磁场，因此，其行为也就像具有南北两极的磁偶极子一样。

永磁棒是具有强极性的磁偶极子。如图 16.31a 所示，如果将两个相同的磁棒北

极对南极放置，两者就会被强大的吸引力吸在一起；反之，如图 16.31b 所示，如果将两者北极对北极（或者南极对南极）靠近放在一起，那么，两个磁棒就会受到强烈的排斥力。磁棒的这两种摆放位置是产生最大作用力的稳定平衡状态。

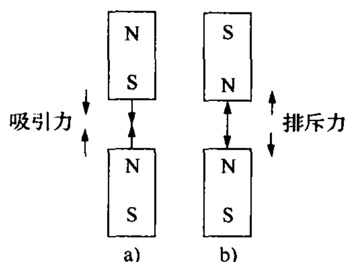


图 16.31 具有南极（S）和北极（N）的两个永磁体在两种不同摆放方式下所产生的吸引力（图 a）和排斥力（图 b）

如果将很多微小的磁偶极子放在强静磁场 B_0 中，那么，如图 16.32 所示，这些磁偶极子要么顺着磁场方向（即平行）排列，要么锁定在逆磁场方向（即反平行）的位置。并且，由于顺磁场排列所需能量较小，因此，多数偶极子会保持这种取向。所有偶极子的偶极矩矢量之和就形成了一个总的净磁化强度，即

$$\hat{\boldsymbol{M}}_0 = \sum_n \hat{\boldsymbol{\mu}}_n \tag{16.54}$$

如图 16.32 中所示。

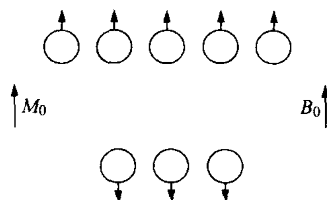


图 16.32 静磁场 B_0 中排列的磁偶极子及其所产生的净磁化强度

从一种自旋状态转变到另一种自旋状态所需的能量为

$$\Delta E = h\nu \tag{16.55}$$

式中 h——普朗克常数 $6.626\times10^{-34}\mathrm{J\cdot s}$，

ν——光子频率。

该式的物理意义是，如果电子吸收了频率为 ν 的光子，它就会跃迁到较高的能级；反过来，如果电子发射了频率为 ν 的光子，它就会降到较低的能级，从而使处于激发态的电子返回到原始平衡状态。与发射电离辐射的 X 射线成像不同，MRI 成像中只发射对人体无损害的光子。

如果要利用外加磁场激发自旋状态的转变，那么，激励磁场的频率应该为多少呢？回到上一节所述的第四种情况，这种情况相当于静磁场中电子电荷绕原子核旋转的情况。根据量子力学理论，自旋电子的旋磁比 γ 等于 q/m。自旋的方向遵循左手螺旋法则，即绕 B_0 顺时针方向。将 γ 代入式（16.53b），可得频率公式为

$$v_L = \frac{\omega}{2\pi} = \frac{\gamma}{2\pi}B_0 = \gamma' B_0 \tag{16.56}$$

这就是所谓的拉莫尔频率，它可以用于计算电子能级跃迁所需的转化能量。注意，该公式表明旋磁比 γ 与角频率相关，其归一化形式 γ' 则与频率相关。式（16.56）是频率与外加磁场相联系的重要公式。对于 MRI 最常用的氢同位素，$\gamma'=42.58\mathrm{MHz/T}$。幸运的是，MRI 成像所需的参数值都在适合检测的范围之内，也就是，频率低于 100MHz，磁场强度为 0.1～4T。

拉莫尔频率就是外加磁场用于诱发自旋状态转变的共振激发频率。MRI 的研究

对象是原子核自旋数为奇数的同位素。例如，顺外加磁场的氢原子^{1}H 的自旋量子数被设定为“1/2”，其磁偶极矩为 $\mu_z=\gamma h\times(+1/2)$。逆外加磁场的^{1}H 的自旋量子数被设定为“−1/2”，其磁偶极矩为 $\mu_z=\gamma h\times(-1/2)$。表 16.2 列出了各种同位素的 γ'值。一般，$\mu_z=\gamma hm$，其中 $m=0$，$\cdots\pm(I-1)$，I 即为表 16.2 所列自旋量子数。

那么，处于不同自旋状态的原子究竟会有多少呢？假设处于低能级和高能级的自旋数量分别为 n_- 和 n_+，则任意时刻处于两个能级的自旋数量之比可用以下 Boltzmann 定律计算

$$n_-/n_+=\exp(-\Delta E/\underline{K}T) \tag{16.57}$$

式中 $\underline{K}$——Boltzmann 常数 1.3805×10^{-23}J/K；

ΔE——两个能级之间的能量差，参见式（16.55）；

T——温度，单位为开尔文（K）。

表 16.2 几种与 MRI 相关的同位素的特性

核　素	自旋量子数	γ'/（MHz/T）	天然同位素丰度（%）	相对于^{1}H 的灵敏度
^{1}H	1/2	42.58	99.99	100
^{13}C	1/2	10.71	1.11	1.6
^{19}F	1/2	40.05	100	3.4
^{23}Na	3/2	11.26	100	9.3
^{31}P	1/2	17.24	100	6.6

例 16.11 设频率为 20MHz，温度为 300K，请计算高能级自旋数量 n_+ 多于低能级自旋数量 n_- 的比例，即 $(n_+-n\ )/n_+$。

解： 已知频率为 20MHz，由式（16.55）可计算高低两个能级之间的能量差为

$$\Delta E=h\nu=6.626\times10^{-34}\text{J}-\text{s}\times20\times10^6/\text{s}=1.325\times10^{-26}\text{J}$$

再由式（16.57）可求得多余的高能级原子数量的比例为

$$(n_+-n_-)/n_+=1-n_-/n_+=1-\exp(-\Delta E/\underline{K}T) \tag{16.58}$$

代入相应的数据，可得

$$(n_+-n_-)/n_+=1-\exp[-1.325\times10^{-26}\text{J}/(1.3805\times10^{-23}\text{J/K}\times300\text{K})]=3.2\times10^{-6}$$

■

这对于整个被测物体而言是一个相当小的数字。实际上，某些同位素具有比较高的出现概率。某种同位素在其元素的所有天然同位素中所占的百分比称为天然丰度，表 16.2 列出了几种同位素的天然丰度。除了同位素丰度以外，在选择医用 MRI 的同位素检测对象时，还有一个需要考虑的重要问题就是所选同位素在人体中的含量有多少。另外，相对检测灵敏度也是需要考虑的因素（参见表 16.2）。值得庆幸的是，人体内的氢原子很丰富，尤其在脂肪和水中特别多。

例 16.12 根据表 16.2 所列数据，说明同位素^{1}H、^{13}C 和^{31}P 是否适用于人体的 MRI 成像。

解： 注意，^{1}H 的天然丰度为 99.98%，而^{13}C 却只有 1.11%，所以，^{13}C 不适合用

于 MRI 成像。另外，^{31}P 虽然很丰富，但其灵敏度很低，很难检测。人体重量的 60% 是水，并且^{1}H 具有很高的灵敏度和丰度，因此，难怪^{1}H 会用于医学 MRI 成像。 ■

16.3.4 拉莫尔进动

要将氢原子磁偶极子激发到可用于成像的自旋状态，必须外加激励磁场。正如 16.3.2 节所述的第四种情况，如果测量线圈垂直于 x 轴放置，那么，由于自旋磁偶极子的方向与 z 轴基本一致，很难测量其产生的磁偶极矩。因此，需要寻找一种方法，能够使磁偶极子倒向 x-y 平面，这样就可以用类似于图 16.29 所示的测量方法检测磁偶极矩。这里需要一个作用力将偶极子推入一种进动状态，使其盘旋运动。其原理与陀螺在重力作用下的转动相似，开始时，陀螺是垂直的，然后，由于重力的作用，陀螺逐渐倾斜，进入一种进动旋转的状态，最后，倒在水平位置。

图 16.33 显示了磁偶极子的这种进动变化过程。开始时，偶极子的净磁化强度矢量$\hat{\boldsymbol{M}}_0$ 是顺静磁场 B_0的 z 轴方向（见图 16.33a）。在 x 轴方向外加另一个以拉莫尔频率交变的磁场 B_1 之后，$\hat{\boldsymbol{M}}_0$ 就会变成以拉莫尔频率绕 z 轴的进动，使$\hat{\boldsymbol{M}}_0$ 与 z 轴之间形成一个 ϕ 角（见图 16.33b）。

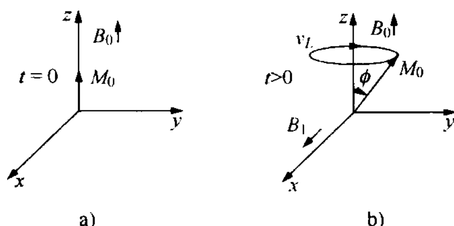


图 16.33　顺静磁场 B_0方向的初始净磁化强度矢量$\boldsymbol{M}_0$（图 a）；x 轴方向外加射频交变磁场 B_1 之后，使得 $\boldsymbol{M}_0$以拉莫尔频率绕 z 轴进动，$\boldsymbol{M}_0$与 z 轴之间形成 ϕ 角（图 b）

为了理解外加射频磁场后$\hat{\boldsymbol{M}}_0$ 的变化，这里有必要引入一个参考坐标系来简化叙述。该坐标系的坐标轴用类似 $\hat{x}'$的方法标记。使用这种参考系后，一方面可以清晰地考察复杂的自旋轨迹和磁化矢量的进动；另一方面可以将拉莫尔频率 ν_L 作为参考频率，简化其他各种频率的检测。

1. 旋转坐标系

上述参考坐标系就是一种旋转坐标系，它相对笛卡尔固定坐标系中的某个点，以拉莫尔频率 ν_L 跟随磁化矢量一起旋转。这种方法与频闪仪的原理相似：相当于将旋转物体看作静止不动的，同步跟随物体的旋转频率，来摄取物体的图像。以频闪仪的频率作为参考，一旦物体的旋转速度变快或者变慢，那么，它与参考频率之差就立即可以测得。同理，在旋转坐标系中，相对于 ν_L 顺时针转动的频率称为正频率，逆时针转动的频率称为负频率。

16.3.2节所述的第三种情况其实是一种测量装置，垂直于 x 轴的测量线圈可以测得永磁体绕 z 轴旋转的信号，它是一个频率为 ω 的正弦信号。如果在 x-y 平面上绕 z 轴旋转的不是永磁体，而是以拉莫尔频率旋转的时变磁化矢量，且矢量的初始位置在 x 轴上，那么

$$\hat{\boldsymbol{B}}_x(t)=\hat{\boldsymbol{x}}\boldsymbol{B}_1\cos\omega t \tag{16.59}$$

现在来看这个磁场矢量在旋转坐标系中会变成怎样。下面用坐标转换方法来求这个新坐标系下旋转磁场新矢量 $\boldsymbol{B}_1'$ 的分量，即

$$B_{x'}=\cos(\omega t)B_x-\sin(\omega t)B_y \tag{16.60a}$$

$$B_{y'}=\sin(\omega t)B_x+\cos(\omega t)B_y \tag{16.60b}$$

将式（16.59）代入，可得

$$B_{x'}=B_1\cos^2\omega t \tag{16.60c}$$

$$B_{y'}=B_1\cos\omega t\sin\omega t \tag{16.60d}$$

利用三角恒等式的半角公式，可以将上式简化为

$$\hat{\boldsymbol{B}}_1'(t)=\hat{\boldsymbol{x}}'B_{x'}+\hat{\boldsymbol{y}}'B_{y'}=\hat{\boldsymbol{x}}'\left(\frac{B_1}{2}-\frac{B_1}{2}\cos2\omega t\right)+\hat{\boldsymbol{y}}'\frac{B_1}{2}\sin2\omega t \tag{16.60e}$$

如果用低通滤波器将两倍拉莫尔频率的成分去除，那么就可得到

$$\hat{\boldsymbol{B}}_1'(t)=\hat{\boldsymbol{x}}'\frac{B_1}{2} \tag{16.60f}$$

不出所料，旋转坐标系中 B_1 的分量看起来是静止不动的，变成了非时变信号。

2. 翻转角和弛豫

在MRI中，磁化矢量位置的控制很重要，图16.34用旋转坐标显示了磁化矢量进动的一种控制方法。首先，初始时的磁化矢量方向与外加静磁场方向一致（见图16.34a）；然后，沿 x' 轴外加一个以拉莫尔频率 ν_L 旋转的射频磁场 B_1，B_1 是一个门控的正弦波短脉冲信号，脉冲持续时间设为 t_p。此脉冲磁场会产生一个作用力，将磁化矢量 $\hat{\boldsymbol{M}}_0$（其初始方向通常沿 z 轴）拉向与 z 轴成 φ 角的进动状态，该 φ 角被称为翻转角，即

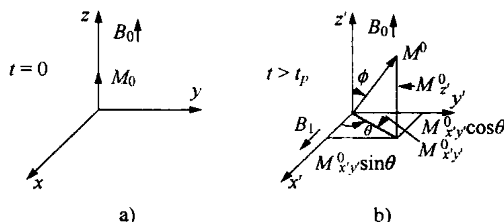


图16.34 在 $t=0$ 时刻的初始磁化矢量方向（图a）；在旋转坐标系中，沿 x' 轴方向的外加射频（拉莫尔频率）磁场 B_1 的作用使得净磁化矢量从 z 轴倾斜了 φ 角（图b）（图中显示了净磁化矢量在各个坐标轴上的投影分量）

$$\varphi=\gamma\left(\frac{B_1}{2}\right)t_p=\gamma B_1' t_p \tag{16.61}$$

如图 16.34b 所示，进动磁化矢量可以分解成旋转系中的各个笛卡尔分量。除了沿 z'轴的分量 $M_{z'}^0$ 以外，还有 x'-y'平面中的总分量，设其幅值为 $M_{x'y'}^0$。虽然 $M_{z'}^0$ 和 $M_{x'y'}^0$ 是同步发生的，但是，单独分开考察这两个分量的复杂时间变化规律，会方便得多。

例 16.13　假设磁场与例 16.10 所述相同，如果要获得 π 弧度的翻转角，请计算所需的外加射频磁场的脉冲宽度。如果线圈产生的磁场强度增加 100 倍，所需的脉冲宽度又是多少？

解：由式（16.56）可知 $\gamma=2\pi\gamma'$；又由例 16.10 可知 $B_1=B_\phi=8\times10^{-6}$；且由表 16.2 得知 $\gamma'=42.58\times10^6$ MHz/T。于是，根据式（16.61），可求得 t_p 为

$$t_p=\frac{\varphi}{2\pi\gamma'(B_1/2)}=\frac{\pi}{2\pi\times42.58\times10^6(/s\cdot T)(4\times10^{-6}T)}=2.936\text{ms}$$

如果磁场强度增加 100 倍，那么，脉冲宽度就可以减小 100 倍，即为 29.36μs。■

如果翻转角选定为 90°（即π/2 弧度），那么，如图 16.35a 所示，在射频脉冲磁场的作用下，磁化矢量就会从初始 z'轴上的 M_0 偏转到 y'轴上，使得 z'方向的垂直分量变为 0。脉冲磁场一旦消失之后，经过一段时间的恢复，垂直分量又会回到初始的 M_0 值。这个恢复过程可以用如下方程描述：

$$M_{z'}(\varphi=\pi/2)=M_0[1-\exp(-t/T_1)] \tag{16.62}$$

式中　T_1——自旋—晶格弛豫时间，由于它与垂直方向相关，故被称为纵向弛豫时间常数；

t——从射频脉冲磁场结束时刻开始计算的时间。

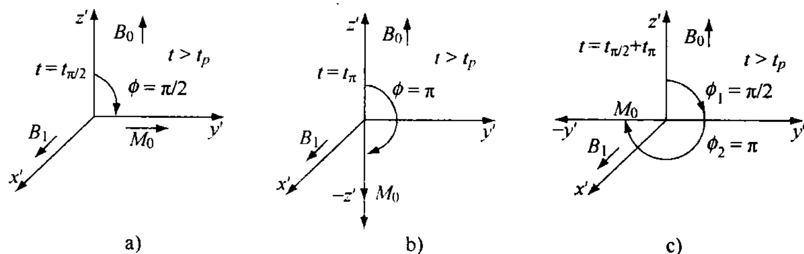


图 16.35　磁化矢量经过π/2 弧度翻转角到达 y'轴（图 a）；磁化矢量经过π弧度翻转角到达 $-z'$轴（图 b）；磁化矢量的连续两次翻转：首先经过π/2 弧度翻转角到达 y'轴，然后再经过π弧度翻转角到达 $-y'$轴（图 c）

图 16.36 中的实线显示了该方程描述的曲线，其弛豫时间与温度和物体的粘性有关，物体的粘性越大，弛豫时间就越长。

如果翻转角选定为 180°（即π弧度），那么，磁化矢量就会从初始 z'轴上的 M_0 转变到 $-z'$轴上，使得 z'方向上的垂直分量变为 $-M_0$。这种反向平行磁化矢量的弛豫方程为

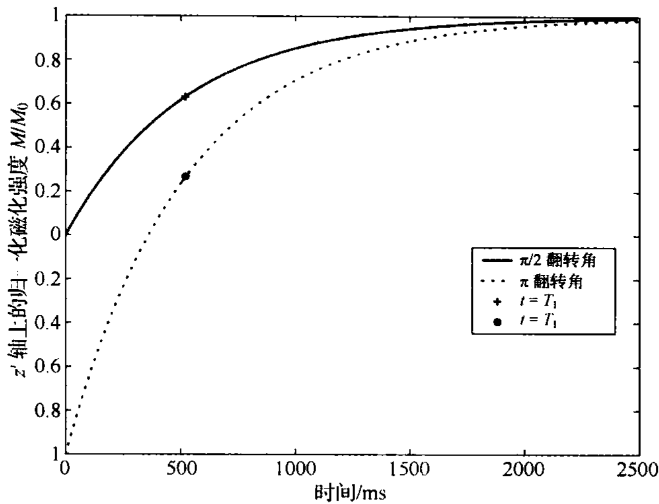


图 16.36 磁化矢量 z'轴分量的恢复曲线（弛豫时间常数为 T_1）

注：图中实线和虚线分别是大脑灰质π/2 和π翻转角的弛豫曲线。

$$M_{z'}(\varphi=\pi)=M_0[1-2\exp(-t/T_1)] \quad (16.63)$$

图 16.35b 显示了这种情况，其弛豫曲线如图 16.36 中的虚线所示。

可以将上述方程推广到磁化矢量 z'轴分量的任意变化值。假设射频脉冲磁场使磁化矢量的 z'轴分量变为 M_z^0，则其弛豫方程为

$$M_{z'}(t)=M_0\left[1-\left(1-\frac{M_{z'}^0}{M_0}\right)\exp(-t/T_1)\right] \quad (16.64)$$

下面来计算 x-y 平面上的磁化矢量分量的变化。我们重新假设翻转角为 $\pi/2$，也就是磁化矢量偏转到 y'轴的情况。设脉冲磁场关闭时刻（$t=0$），x'-y'平面上的磁化矢量初始值为 $M_{x'y'}^0$，那么，脉冲磁场消失后 $M_{x'y'}^0$ 立即会逐渐衰减，回到 0 值，从而使磁化矢量恢复到原来的纯垂直分量 M_0。$M_{x'y'}^0$ 的弛豫方程为

$$M_{x'y'}(\varphi=\pi/2)=M_{x'y'}^0[\exp(-t/T_2)] \quad (16.65)$$

式中 T_2——自旋—自旋弛豫时间。

此方程表示的弛豫曲线如图 16.37 所示。由于磁化矢量的这个横向分量可以移动和旋转，因此，它与生物组织等物质的相互作用就与前述垂直分量不同。如图 16.38 所示，随着磁化矢量扫过的区域逐渐增大，原子核系统发生相散，使磁化矢量横向分量的净总和逐渐趋于零。

T_2 与 T_1 的意义不同。T_2 对分子之间的相互作用和外加磁场 B_0的非均匀性较敏感。该弛豫常数还有一种比较切合实际应用的定义，它包括了磁场变化 ΔB_0的效应，即

$$\frac{1}{T_2^*}=\frac{1}{T_2}+\frac{\gamma\Delta B_0}{2} \quad (16.66)$$

比较图 16.36 和图 16.37 可见，T_2 显然比 T_1 短，而 T_2^* 则是三者中最短的。表 16.3

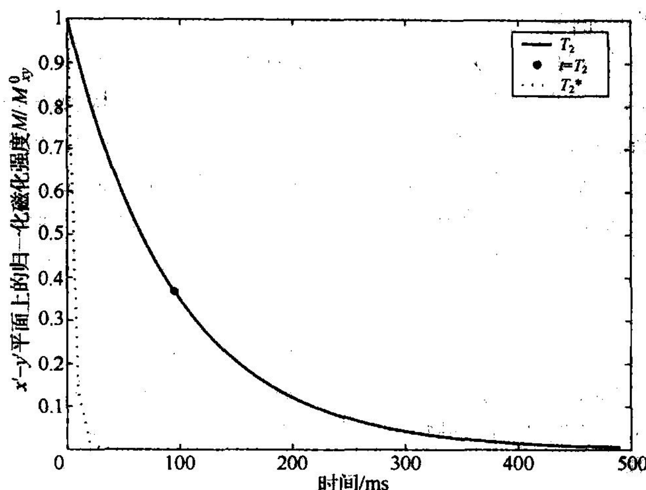


图 16.37 大脑灰质π/2 翻转角的磁化矢量在 x'-y'平面上的分量的弛豫曲线（弛豫时间常数为 T_2，$T_2^*=5$ms）

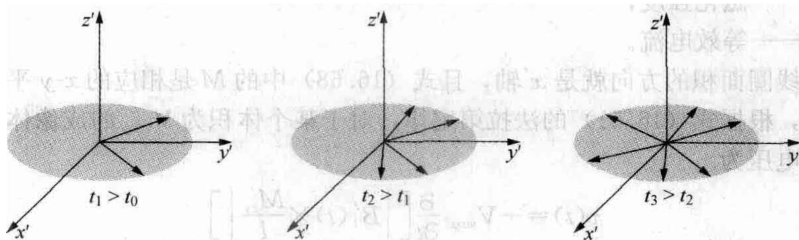


图 16.38 π/2 翻转角发生之后自旋一自旋弛豫的相散时间序列（弛豫时间常数为 T_2）

列出了各种人体组织常用的 T_1 和 T_2 时间常数。

例 16.14 在产生 π/2 翻转角之后，当磁化矢量的水平分量衰减为初始值的 1/e 时，脂肪组织中与 T_1 相关的磁化矢量的垂直分量恢复了百分之几？

解：由表 16.3 可知，脂肪的 $T_1=160$ms，且 $T_2=100$ms。T_2 就是磁化矢量水平分量衰减到 1/e 初始值所需的时间。根据式（16.62），此时垂直分量的恢复值为

$$1-\exp(-T_2/T_1)=0.465=46.5\%$$ ■

表 16.3 几种人体组织的常用 T_1 和 T_2 值

组　织	T_1/ms	T_2/ms
脑脊液	2000	1000
脂肪	160	100
灰质	520	95
恶性肿瘤	800	200
水肿或梗塞	600	150
白质	380	85

3. MRI 检测的信号

至此我们已经讲述了有关核磁共振激发和衰减的背景知识，接下来就可以介绍 MRI 检测的信号。测量装置的原理如图 16.33a 所示，检测线圈平面与 x 轴垂直。引发 90°翻转角的$\pi/2$ 射频磁场激励脉冲是一个脉宽为 t_p 的正弦波短脉冲信号，正弦波的频率就是拉莫尔频率，t_p 值由式（16.61）决定。由于激励脉冲的持续时间比任何恢复时间常数和衰减时间常数都要长得多，因此，可以用一个按指数规律变化的正弦波信号来简化分析过程。x-y 平面上的横向磁化矢量可以写为

$$\hat{\boldsymbol{M}}_{xy}(t)=M_{xy}^{0}\mathrm{e}^{-|t|/T_2}\mathrm{e}^{\mathrm{j}\omega t}(\hat{\boldsymbol{x}}+\hat{\boldsymbol{y}})\qquad t<t_p \tag{16.67}$$

该磁场激励的作用是改变氢原子中电子的自旋状态并诱发共振。利用法拉第定律可以检测本章 16.3.2 节所述第三种情况下的总的净磁化响应。假设直径为 d 的测量线圈的面积为 S，则

$$S=\frac{M}{I} \tag{16.68}$$

式中 M——磁化强度；

I——等效电流。

如果线圈面积的方向就是 x'轴，且式（16.68）中的 M 是相应的 x-y 平面上的分量，那么，根据式（16.51）的法拉第定律，对于某个体积为 V_{voxel} 的成像体素，其产生的感应电压为

$$v(t)=-V_{\text{voxel}}\frac{\partial}{\partial t}\left[\left|B_1'(t)\times\frac{M_{xy}}{I}\right|\right]$$

$$v(t)=-V_{\text{voxel}}\frac{\partial}{\partial t}\left[\left|\hat{\boldsymbol{B}}_1'(t)\times\frac{\hat{\boldsymbol{M}}_{xy}}{I}\right|\right]=-V_{\text{voxel}}\frac{\partial}{\partial t}\left[|B_1'(t)|\left|\frac{M_{xy}}{I}\right|\sin\theta\right] \tag{16.69a}$$

由式（16.67）可得 M_{xy}，且由式（16.60f）可得 $\hat{\boldsymbol{B}}_1'(t)$。并且，$\hat{\boldsymbol{B}}_1'$与 $\hat{\boldsymbol{M}}_{xy}$之间的夹角 $\theta=90°$，因此，上式所表示的电压就变为

$$v(t)=\left(\frac{B_1}{2I}\right)\cdot(\mathrm{j}\omega V_{\text{voxel}}M_{xy}^{0}\mathrm{e}^{-|t|/T_2}\mathrm{e}^{\mathrm{j}\omega t}) \tag{16.69b}$$

式中 V_{voxel}——图像体素的三维体积。

该磁共振电压信号 $v(t)$ 被称为自由感应衰减（Free Induction Decay，FID）信号，图 16.39 显示了此信号的实部。

但是，FID 信号并不适合用于 MRI 检测，主要是因为它受磁场变化的影响太大，而磁场决定了原子核自旋状态的响应信号。另外，FID 信号的衰减时间常数很小，信号衰减太快，难以直接测量。如图 16.38 所示，随着原子核进动的空间位置的变化，其自旋状态要经历非均匀磁场的作用，从而导致各个原子核磁化矢量角频率之间的差异。

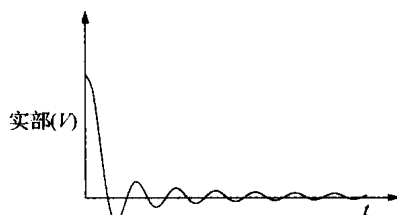


图 16.39 随时间变化的自由感应衰减信号（信号从 $t=0$ 开始）

由式（16.56）可知，某些原子核的进动速度要比其他原子核快。因此，使原子核系统迅速相散，测不到磁共振信号。

自旋回波（Spin Echo，SE）法是一种巧妙的间接测量方法。如图 16.35c 所示，先施加一个 $\pi/2$ 激励脉冲，使磁化矢量偏转到 y 轴；然后，激励脉冲结束后经过某个 $T_E/2$ 时间，当自旋矢量发生明显相散但还未完全失相时，再施加第二个 π 激励脉冲，其幅值为原 $\pi/2$ 脉冲的两倍。如图 16.40a 所示，该 π 激励脉冲可以将所有不同相位的原子核翻转 180°，从而使各个磁化矢量又向反方向进动。由于各自的速度保持不变，再经过 $T_E/2$ 时间后，所有原子核又重新聚集到 y 轴上，变成同相位。于是，就会在 T_E 时刻再次出现磁共振信号的一个波峰，然后再衰减，这就形成了图 16.40b 所示的自旋回波信号。这种方法的妙处就在于它抵消了磁场不均匀性的影响，实现了人体组织 T_2 响应信号的准确测量。与 FID 信号的推导方式相似，自旋回波信号的方程如下：

$$v(t)=\mathrm{j}(B_1/2I)\omega V_{\mathrm{voxel}}M_{xy}^0\mathrm{e}^{-|t/T_2|}\mathrm{e}^{\mathrm{j}\omega t} \tag{16.70}$$

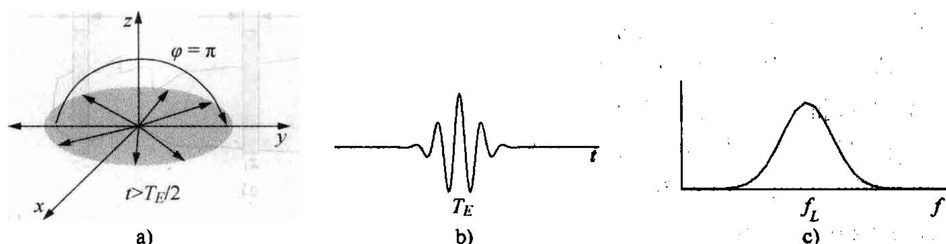


图 16.40　π 激励脉冲引起各个原子核磁化矢量旋转方向的 180°翻转，使它们重新向同相位运动（图 a）；自旋回波信号的时间函数（图 b）；自旋回波信号的幅值频谱（图 c）

利用以下两个傅里叶变换，可以求得自旋回波信号的频谱。一个变换是

$$\Im[\exp(-|t|)]=\frac{2}{1+(2\pi f)^2} \tag{16.71a}$$

另一个变换是

$$\Im(\mathrm{e}^{\mathrm{j}\omega_L t})=\delta(f-f_L) \tag{16.71b}$$

注意，式（16.70）中的 ω 是个常数，且 $\omega=\omega_L=2\pi f_L$，其物理意义相当于拉莫尔频率 v_L。

再根据式（16.4a）所述的傅里叶比例变换性质，于是，式（16.70）的回波信号的频谱就成为

$$V(f)=\Im\left(\frac{\mathrm{j}B_1M_{xy}^0\omega_L V_{\mathrm{oxel}}\mathrm{e}^{-|t/T_2|}\mathrm{e}^{\mathrm{j}\omega_L t}}{2I}\right)=\frac{\mathrm{j}B_1M_{xy}^0\omega_L V_{\mathrm{voxel}}T_2}{2I}\left\{\frac{1}{1+[2\pi T_2(f-f_L)]^2}\right\} \tag{16.71c}$$

该频谱的幅值如图 16.40c 所示。

如果定义 $S(f)$ 函数为

$$S(f)=\frac{\mathrm{j}B_1M_{xy}^0\omega_LV_{\mathrm{voxel}}T_2}{2I}\left\{\frac{1}{1+[2\pi T_2f]^2}\right\} \qquad (16.71\mathrm{d})$$

那么式（16.71c）就变为

$$V(f)=S(f-f_n) \qquad (16.71\mathrm{e})$$

此处 $f_n=f_L$。一般，下标 n 表示 x_n 处的拉莫尔激励频率。下一节就将介绍这种对应关系的由来。

16.3.5 磁共振成像设备

为了根据测量信号重建被测物体的 MRI 图像，必须建立一种被测信号的空间定位方法。MRI 成像中，笛卡尔三维坐标系中的 3 个坐标轴各自有一种被测数据与坐标位置之间的编码方法，所以，下面将分别介绍 z 轴、x 轴和 y 轴的定位。三维图像由许多离散化的微小立方体组成，这些立方体被称为体素（voxel），每个体素的体积为 $V_{\mathrm{voxel}}=\Delta x\Delta y\Delta z$。

创建图像的第一步是要确定所需成像的断层平面，如图 16.41 所示，它是 z 轴上厚度为一个体素的切面。MRI 检测时，病人被推入一个能够产生强静磁场 B_0 的巨型超导磁体中。回顾一下 16.3.2 节所述的第一种和第二种情况：导线或线圈中的电流可以产生磁场。MRI 的磁体中就是安装了多个线圈，这些线圈可以产生一个 z 轴方向的电控线性梯度磁场，即

$$\vec{B}_z=\hat{z}(B_0+G_zz) \qquad (16.72\mathrm{a})$$

式中 G_z——磁场梯度常数，单位为 T/m。

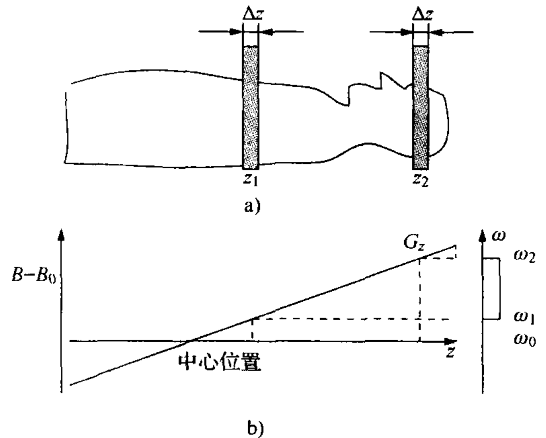


图 16.41 轴向成像平面位置的侧视图（图 a）；在 z 轴梯度磁场上选中的两个成像断层的位置（图 b）

注：图中右侧的垂直坐标显示了射频激励脉冲选择拉莫尔频率的范围。为了能够看得清楚，此图没有按比例绘制。

由式（16.56）可知，该梯度磁场可以在 z 轴的每个位置上产生一个特定的拉莫尔频率，即

$$\omega=\gamma B_z=\gamma B_0+\gamma G_zz \qquad (16.72\mathrm{b})$$

$z=0$ 处被设定为中心位置，该中心位置的拉莫尔频率为 $\omega_0=\gamma B_0$，被称为中心频率，它可以通过电子处理方法从 ω 中减去。

如图 16.41 所示，如果外加激励磁场的频率在合适的范围内，那么，就可以选中某个特定的成像平面。为了获得较平稳的激励磁场，时域上的射频激励脉冲的正弦波包络线形状需要进行适当的调制。sinc 函数形状的脉冲具有平坦的方波

频谱，也就是，其傅里叶变换是一个方波函数。但是，理论上，这种函数在时域上是无限长的。图16.42显示了sinc脉冲及其相应的磁场频谱形状。时域sinc脉冲函数的傅里叶变换F可以用带宽为Δf的方波函数表示如下

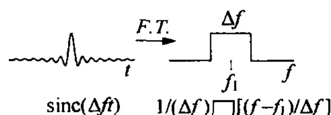


图16.42 具有选定拉莫尔中心频率的Sinc调制激励脉冲及其傅里叶变换（其频谱可以表示外加激励磁场的形状）

$$\mathfrak{F}[\operatorname{sinc}(\Delta f t)\exp(\mathrm{j}2\pi f_1 t)]=(1/\Delta f)\Pi[(f-f_1)/\Delta f] \qquad (16.73)$$

作为一种折中方法，实际使用的是直接截断或者加窗截断的有限长sinc脉冲，图16.40b和c所示就是加了锥形递减窗的sinc脉冲波形及其幅值频谱。

例16.15 假设磁场梯度常数$G_z=0.352\mathrm{T/m}$。如果要获得10mm层厚的z轴断层图像，请问激励磁场的频谱带宽应该为多少？如果静磁场$B_0=1\mathrm{T}$，离开中心位置10cm和20cm处的断层成像分别需要使用中心频率为多少的激励磁场？

解： 将$\gamma=2\pi\gamma'$代入式（16.72b），就可以计算如下频率差（即频谱带宽），它与断层厚度成正比，即

$$\Delta f=42.58(\mathrm{MHz/T})\times 0.352(\mathrm{T/m})\times 0.10\mathrm{m}=0.15\mathrm{MHz}$$

同样，利用此公式也可以计算偏离中心位置10cm处的断层所需的中心频率为$f_1=42.58(\mathrm{MHz})[1\mathrm{T}+0.352(\mathrm{T/m})\times 0.1\mathrm{m}]=44.08\mathrm{MHz}$，同理，20cm处所需的中心频率为$f_2=45.58\mathrm{MHz}$。■

用梯度磁场确定了扫描断层在z轴上的位置之后，如图16.43所示，下一步就是要确定x-y平面上的位置。同理，x轴上的空间定位也利用外加梯度磁场来确定，即

$$\omega=\gamma B_x=\gamma B_0+\gamma G_x x \qquad (16.74\mathrm{a})$$

当$B=B_0$时，也有$\omega_{\mathrm{mix}}=\omega_0$。消去$\omega_0$之后，频率差与位置$x$成正比：

$$\Delta\omega=\omega-\omega_{\mathrm{mix}}=\gamma G_x x \qquad (16.74\mathrm{b})$$

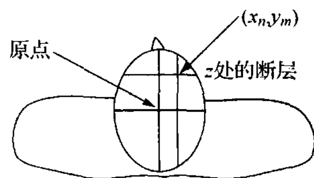


图16.43 在z轴上位置z处的某个层厚为Δz的x-y断层平面（参见图16.41所示断层位置）

根据式（16.72b），前述选择z轴上某个断层位置时使用了特定的激励磁场频率。虽然选定x轴上位置时也应用同样的原理，但实现方式有所不同。这里，被测磁共振信号的频率随x轴上的距离变化，如图16.44所示，是一个线性梯度磁场信号。沿x扫描线的一系列共振激发信号组成了整个测量

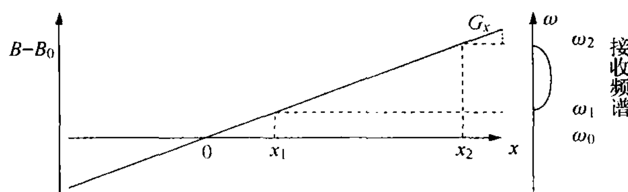


图16.44 x轴上扫描位置的频率编码方法（图中显示了x_1和x_2两个位置）

信号，信号的频率变化编码了 x 轴的位置信息。

超声成像中所检测的回波信号的时间延迟与超声传播途径中不同深度距离上遇到的反射物相关。与这种超声成像原理不同，不同空间位置上的 MRI 信号是由频率编码的，它们都叠加在相同的时间点上。在频域上为

$$V(f)=\sum_{n}S(f-f_n) \tag{16.75}$$

下一节会介绍此 x 轴位置信号的解码方法。

虽然 MRI 系统也曾经采用过外加 y 轴方向的线性梯度磁场来定位 y 轴坐标，但是，现在比较常用的不是这种频率编码方法，而是另一种相位编码方法。其依据是式（16.61）表示的翻转角理论，此处变为

$$\varphi(y)=\omega t=-(\gamma B_0 t+\gamma G_y t_p y) \tag{16.76}$$

其中 $y=y'+y_0$，y_0是中心位置。根据此翻转角理论，具有特定脉宽 t_p 的射频磁场可以使质子的磁化矢量旋转一个所需的角度。由式（16.76）可知，改变脉宽 t_p 或者改变梯度常数 G_y 都可以引起磁化矢量相位的变化。

如图 16.45 所示，比较实用的方法是通过改变 y 轴梯度线圈的电压幅值来调节梯度常数 G_y 的大小和方向，而不是改变脉宽 t_p。为了获得整个 x-y 平面的成像信息，一般使用 128 或 256 个不同的激励信号扫描 y 轴，每个激励信号对应于一个特定的 y 轴位置。图 16.46 显示了顺序排列的几条扫描线的控制脉冲和激励信号。

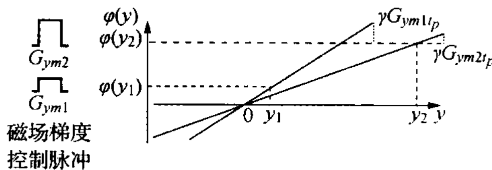


图 16.45 y 轴位置的相位编码方法（图中显示了 y_1和 y_2两个位置）

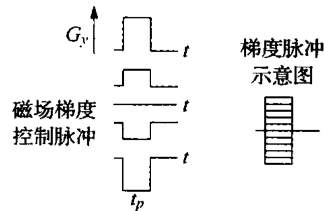


图 16.46 几条依次排列的扫描线脉冲序列

注：请注意 G_y 脉冲的幅值变化过程，一般 G_y 的变化范围为 G_{ymax}～$-G_{ymax}$。

16.3.6 磁共振成像方法

前面已经将 MRI 成像所需的几乎所有基础知识都介绍过了。现在，我们要逐步总结各种不同的控制和检测方法，并将它们整合起来，用于阐述 MRI 成像的实际过程。为了简洁明了，某些过程将适当简化。图 16.43 所示是颅脑扫描的一个 x-y 成像平面，其中包含了中心位置，也就是磁场坐标系的原点。下面就介绍此平面中某个特定位置上的净磁化信号转换成磁场强度、被测信号以及最终转换为图像信号的过程。

图 16.47 显示了 MRI 测量的脉冲时序：

首先是两个射频脉冲信号 $\pi/2$ 脉冲和 π 脉冲。正如 16.3.4 节所述，这两个射频激励脉冲迫使磁化矢量进动到 x-y 平面，从而建立自旋回波信号。z 轴的梯度脉冲信

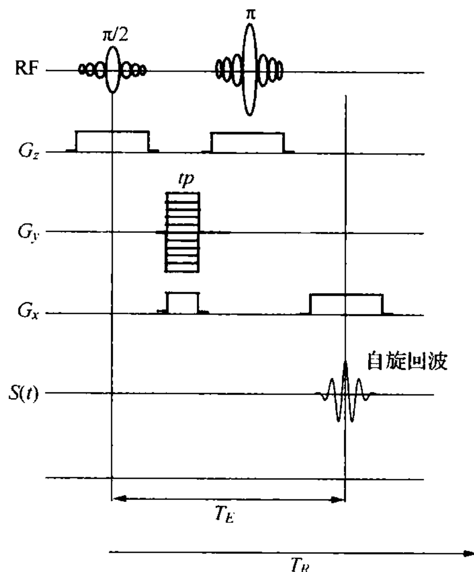


图 16.47 MRI 的基本射频激励信号以及用于自旋回波信号 $S(t)$ 检测的梯度脉冲序列 G_z、G_y 和 G_x。T_E 为回波时间（T_R 为重复周期，也就是两个基本时序之间的时间间隔）

号 G_z 与射频脉冲同步发生，用于接通梯度线圈，使线圈发射选定频率的射频脉冲，进行 16.3.5 节所述的 x-y 成像平面在 z 轴上的定位。

然后，在原子核进动期间和自旋回波形成过程中，发射 x 轴方向的梯度脉冲。通过该梯度磁场的作用，将原子核自旋信号在 x 轴上的位置信息转变成不同的频率信号。

最后，再用梯度脉冲激励 y 轴线圈，并通过脉冲的幅值实现 y 轴位置的相位编码。

此后，在 $t=T_E$ 时刻检测自旋回波信号。上述基本时序的重复时间间隔为 T_R。

利用傅里叶变换可以将采集到的一系列信号序列与被测物体的磁共振图像（即净磁化矢量的分布）关联起来。假设成像断层的 z 轴位置已经选定，那么，物体的图像就是坐标位置（x，y）的函数，需要使用二维傅里叶变换。如果能够先将 x 和 y 两个坐标分开考虑，单独进行一维傅里叶变换，然后再将它们合成二维变换，那就比较简单了。下面分析成像断层上（x_n，y_m）点处的净磁化分布信号。

为了便于理解，我们先回顾一下 16.1.1 节所述的冲激函数及其傅里叶变换的性质，它们对于 MRI 图像的重建技术具有重要的作用。冲激函数是一个具有特殊性质的广义函数，正如 16.1.1 节所述，它对被积函数具有采样作用，即

$$\int_{-\infty}^{\infty} \delta(f-f_0)R(f)\mathrm{d}f = R(f_0) \tag{16.77a}$$

冲激函数的傅里叶反变换是一个指数函数，即

$$g(t) = \int_{-\infty}^{\infty} \delta(f-f_0)\mathrm{e}^{\mathrm{i}2\pi ft}\mathrm{d}f = \mathrm{e}^{\mathrm{i}2\pi f_0 t} \tag{16.77b}$$

而该指数函数的傅里叶变换就是

$$G(f)=\int_{-\infty}^{\infty}(e^{i2\pi f t_0})e^{-i2\pi f t}dt=\int_{-\infty}^{\infty}e^{-i2\pi t(f-f_0)}dt=\delta(f-f_0) \tag{16.77c}$$

如果函数 $w(t)$ 的傅里叶变换为 $W(f)$，那么

$$\int_{-\infty}^{\infty}[w(t)e^{i2\pi f_0 t}]e^{-i2\pi f t}dt=\int_{-\infty}^{\infty}w(t)e^{-i2\pi t(f-f_0)}dt=W(f-f_0) \tag{16.77d}$$

现在，我们先从 y 轴开始分析 MRI 图像的建立方法。MRI 的 y 位置信息是一种相位编码，并且由式（16.76）决定。消去净磁场 B_0 产生的相位之后，式（16.76）可以简化为

$$\varphi(y)=-\gamma G_y t_p y \tag{16.78a}$$

与 y 轴上特定位置 y_m 相对应的相位就是

$$\varphi_m(y_m)=-(\gamma G_{y,m} t_p y_m) \tag{16.78b}$$

此相位是位置 y_m 和特定的磁场梯度系数 $G_{y,m}$的函数。相应的被编码时间信号可以描述为

$$s_m(0,\ G_y)=e^{i\gamma t_p y_m G_y} \tag{16.78c}$$

如图 16.45 所示，梯度系数 G_y 是确定相位的变量。以 G_y 为变量进行傅里叶变换时，傅里叶域中的新变量用 u 表示。我们知道，时间 t 与频率 f 这对傅里叶变换变量的单位之间存在倒数关系。同理，u 的单位也是 G_y 单位的倒数，即 m/T。G_y 和 u 这两个变量可用于构建特定的相位项，并且，还可以通过傅里叶变换与 y_m 联系起来。下面我们就来分析这个关系。设傅里叶变换

$$S_m(0,\ u)=\int_{-\infty}^{\infty}(e^{i2\pi u_m G_y})e^{-i2\pi u G_y}dG_y=\int_{-\infty}^{\infty}e^{-i2\pi(u-u_m)G_y}dG_y=\delta(u-u_m) \tag{16.79}$$

这里 u 出现在被积函数第二个指数项的参数中。根据式（16.77c）和式（16.78a），此第二个指数项的整个参数就是相位，因此，由以下等式可以求得 u 的显式函数，即

$$2\pi u G_y=-\gamma t_p y G_y \tag{16.80a}$$

$$u=-\gamma t_p y/(2\pi) \tag{16.80b}$$

可见，该相位编码的傅里叶变换是根据冲激函数的性质得到的。将式（16.80b）代入式（16.79），进行变量代换，可得

$$S_m(0,\ u)=\delta(u-u_m)=\delta[(-\gamma t_p/(2\pi))(y-y_m)] \tag{16.81a}$$

再利用 16.1.1 节所述的冲激函数的性质 $\delta(au)=\delta(u)/|a|$，就有

$$S_m(0,\ y)=[2\pi/(\gamma t_p)]\delta(y-y_m) \tag{16.81b}$$

这是 MRI 的一个关键方程，其推导过程揭示了磁共振成像的重要原理之一。首先，根据式（16.78b），某处的梯度系数 $G_{y,m}$将 y 轴上的位点 y_m 编码成为相位。其次，在傅里叶变换域，该相位使得磁共振信号频谱的中心变为 u，而 u 值与相应的净磁化矢量的位置相关。最后，通过简单的比例变换，就可以将信号频谱转换到 y 轴空间位置上（见图 16.48a）。因此，该过程将尺度变化了的磁共振信号幅值频谱映射到了 y 轴上。

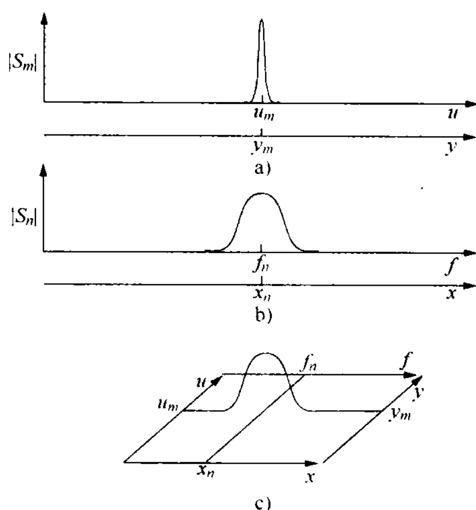


图 16.48　同时以 y 轴和 u 轴为横坐标作出的相位编码磁共振信号的幅值频谱曲线，其中 y 轴和 u 轴经过比例变换（图 a）；同时以 x 轴和 f 轴为横坐标作出的频率编码磁共振信号的幅值频谱曲线，其中 x 轴和 f 轴经过比例变换（图 b）；xy 平面（即 uf 平面）上的频谱曲线（图 c）（为了清晰，图中尺度被放大）

以上是 y 轴的分析结果，下面我们来建立 x 轴上的傅里叶变换关系。假设单个自旋回波信号如图 16.40b 所示，它位于 $y=0$ 位置，因此 $G_y=0$。由式（16.70）可知，经过时间延迟 T_E，并经过特定时间常数 T_{2n} 的衰减之后（T_{2n} 与位置 x_n 相关），该自旋回波信号变为

$$S_n(t,\ 0)=2S_0\mathrm{e}^{-|(t-T_E)/T_{2n}|}\mathrm{e}^{\mathrm{j}\omega_n(t-T_E)} \tag{16.82}$$

式中，S 是 t 和 G_y 的函数，且 $G_y=0$，S_0 为常数。与位置 x_n 相关的拉莫尔频率 ω_n 为

$$\omega_n=\gamma G_x x_n \tag{16.83}$$

注意，中心位置的 ω_0 已被消去。

式（16.82）的一维傅里叶变换是

$$S_n(f,\ 0)=\int_{-\infty}^{\infty}s(t,\ 0)\mathrm{e}^{-\mathrm{i}2\pi ft}\mathrm{d}t=\int_{-\infty}^{\infty}[S_0\mathrm{e}^{-|t-T_E|/T_{2n}}][\mathrm{e}^{-\mathrm{i}2\pi f_nT_E}\mathrm{e}^{-\mathrm{i}2\pi(f-f_n)t}]\mathrm{d}t \tag{16.84a}$$

将此式与式（16.77d）进行对比，可见，此式其实可以看作位于 f_n 处的冲激函数的傅里叶变换。这就意味着左方括号内函数的傅里叶变换是以 f_n 为中心的。为了说明这个结果，有必要确定该方括号中函数的傅里叶变换是否具有式（16.71a）和式（16.71d）的形式，即

$$W_n(f,\ 0)=\int_{-\infty}^{\infty}w_n(t,\ 0)\mathrm{e}^{-\mathrm{i}2\pi ft}\mathrm{d}t=\int_{-\infty}^{\infty}[S_0\mathrm{e}^{-|t-T_E|/T_{2n}}]\mathrm{e}^{-\mathrm{i}2\pi ft}\mathrm{d}t=\frac{S_0}{1+(2\pi fT_{2n})^2} \tag{16.84b}$$

S_0 为常数。因此，$S_n(f,\ 0)$ 就可以表示为

$$S_n(f, 0)=\mathrm{e}^{-\mathrm{i}2\pi f_n T_E} W_n(f-f_n) \tag{16.84c}$$

根据式（16.83），对上式中的频率变量进行代换，就得到

$$S_n(x, 0)=\mathrm{e}^{-\mathrm{i}\gamma G_x x_n T_E} W_n\left[\frac{\gamma G_x}{2\pi}(x-x_n)\right] \tag{16.84d}$$

这也是 MRI 的一个关键方程，其推导过程揭示了磁共振成像中的另一个重要原理。首先，局部的受激磁共振信号将位置信息编码成为频率 f_n，作为时间信号相位的一部分。其次，在傅里叶变换域，该相位使得磁共振信号的频谱中心变为频率 f_n，而 f_n 与净磁化强度的位置相关。最后，经过简单的比例变换，磁共振信号的频谱就被转换到 x 轴的空间位置上（见图 16.48b）。该过程将尺度变化了的磁共振信号幅值频谱映射到 x 轴上。

以上分别推导了 y 轴和 x 轴的一维信号，为了将它们组合成二维信号，可以将两个一维傅里叶反变换结合在一起，得到如下二维函数：

$$s(t, G_y)=\mathfrak{F}\mathfrak{F}^{-1}[I(f, u)]=\iint I(f, u)\mathrm{e}^{\mathrm{i}2\pi ft}\mathrm{e}^{\mathrm{i}2\pi uG_y}\mathrm{d}f\mathrm{d}u \tag{16.85}$$

为了根据（x_n，y_m）位置上的测量信号计算磁化矢量的强度分布 I，再对上式求二维傅里叶变换，即

$$I(f, u)=S_{mn}(f, u)=\mathfrak{F}\mathfrak{F}[S_{mn}(t, G_y)]=\iint S_{mn}(t, G_y)\mathrm{e}^{-\mathrm{i}2\pi ft}\mathrm{e}^{-\mathrm{i}2\pi uG_y}\mathrm{d}t\mathrm{d}G_y \tag{16.86}$$

这里，函数 $s_{mn}(t, G_y)$ 是两个一维信号的简单相乘，即

$$s_{mn}(t, G_y)=\left[S_0\mathrm{e}^{-|(t-T_E)/T_{2mn}|}\mathrm{e}^{\mathrm{i}\omega_n(t-T_E)}\right]\mathrm{e}^{\mathrm{i}2\pi y_m G_y} \tag{16.87}$$

因此，将此式代入式（16.86），可得二维傅里叶变换为

$$s_{mn}(x, y)=\left\{\mathrm{e}^{-\mathrm{i}\gamma G_x x_n T_E} W_{mn}\left[\frac{\gamma G_x}{2\pi}(x-x_n)\right]\right\}[2\pi/(\gamma t_p)]\delta(y-y_m) \tag{16.88}$$

如图 16.48c 所示，在这种情况下，两个一维函数的交点就构成了（x_n，y_m）位点上的尺度变换了的频谱。u_m 的作用是通过尺度的比例变换确定频谱在 y 轴上的位置。在 MRI 图像中，频谱 $S_{m,n}$ 被映射为图像的灰度级，白色表示 z 断层上的最大值。由此可见，MRI 图像显示的并不是测量的磁共振时间信号，而是时间信号的频谱幅值。通过上述这些编码方法，这些频谱幅值被映射到相应的空间位置上。

MRI 实际测量的感应电压信号是被扫描断层 z 平面上所有位点产生的净磁化矢量信号之和，即

$$V(t, G_y)=\sum_{m,n} s_{m,n}(t, G_y) \tag{16.89}$$

这些测量信号并不是式（16.87）和式（16.88）推导出来的理想形式，而是反映被测人体组织特性和组成结构的信号。

k 空间（k-space）方法是分析 MRI 采样数据及其傅里叶变换的一种简单方法。它用 k 空间描述原始测量信号，并用（x，y）空间描述相应的傅里叶变换信号。如

下所述，通过简单的变量代换，就可以实现这种描述。先从包含了重要变量的式（16.83）和式（16.80a）开始，将这两个公式变换成如下两个新公式：

$$2\pi f_n t = k_x x_n \tag{16.90a}$$

$$2\pi u G_y = k_y y_m \tag{16.90b}$$

根据前述定义，很容易看出

$$k_x = \gamma G_x t \tag{16.91a}$$

$$k_y = -\gamma G_y t_p \tag{16.91b}$$

即

$$t = k_x / \gamma G_x \tag{16.92a}$$

$$G_y = -k_y / \gamma t_p \tag{16.92b}$$

再将这两个 t 和 G_y 的公式代入式（16.85），可得

$$s(k_x/\gamma G_x,\ -k_y/\gamma t_p) = \iint \left\{ \left(\frac{\gamma G_x}{2\pi}\right)\left(\frac{-\gamma t_p}{2\pi}\right) I\left[\left(\frac{\gamma G_x}{2\pi}\right)x,\ \left(\frac{-\gamma t_p}{2\pi}\right)y\right] \right\} e^{ik_x x} e^{ik_y y} \mathrm{d}x\,\mathrm{d}y \tag{16.93}$$

用新变量 $\tilde{s}$ 和 $\tilde{I}$ 重新定义上式中的 s 和 I 函数，可得

$$\tilde{s}(k_x,\ k_y) = \iint \tilde{I}(x,\ y) e^{ik_x x} e^{ik_y y} \mathrm{d}x\,\mathrm{d}y \tag{16.94}$$

其中 $\tilde{I}(x,\ y)$ 等于式（16.93）指数前面大括号中的项。同理，式（16.86）可以变为

$$\tilde{I}(x,\ y) = \iint \tilde{s}(k_x,\ k_y) e^{-ik_x x} e^{-ik_y y} \mathrm{d}k_x \mathrm{d}k_y \tag{16.95}$$

这样，前面所述的分析过程就都可以用 xy 空间和 k 空间的变量来完成。

16.3.7 磁共振成像系统

图 16.49 显示了 MRI 系统的主要组成部分。其中，巨型超导磁体用于产生均匀的静磁场 B_0。临床使用的静磁场强度一般在 0.15～2T 的范围内，但也有使用 3～5T 磁场强度的 MRI 系统。主磁体内还安装有用于产生 x、y 和 z 轴梯度磁场的成对表面线圈。这些线圈产生的磁场强度通常只有主静磁场的百分之几，它们用于磁自旋共振信号的定位。被测人体所产生磁共振信号被射频发射/接收线圈检测，该线圈有时制作成靠近人体放置的表面线圈。计算机（即 CPU）控制射频脉冲发生器和时序发生器产生射频信号。切换开关将这些射频脉冲送给射频线圈，同时又将接收到的射频信号送给放大

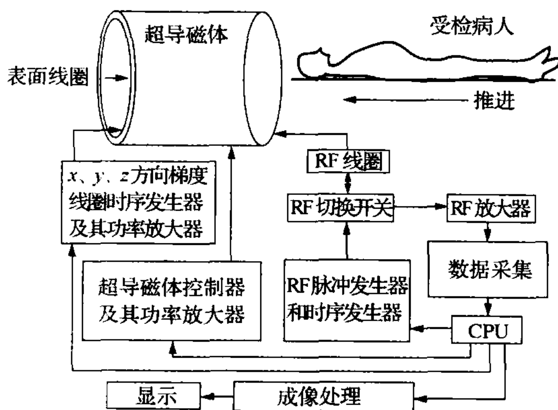


图 16.49 MRI 系统的结构框图

器。然后，接收信号被数据采集系统采样，并由计算机处理后经过傅里叶变换生成图像。新型 MRI 系统的灵活性很大，用户可以通过改变各个磁场梯度和脉冲时序来获得不同的成像效果和不同的加权像。图像采集时间一般为 50～100ms 数量级，不过，现在已经有更快的采集速度。

MRI 检测时，可以通过调节图 16.47 所示的基本脉冲时序，来强调不同的信号特征。例如，如果缩短脉冲重复时间间隔（TR）和自旋回波时间（TE），就可以获得图 16.50b 所示的常见的 T_1 加权像（T_{1W}）。典型的取值是 TE=12ms，TR=300ms。如果延长 TR 和 TE，并使 TE 前后的延时变长，那么，就可以获得图 16.50a 所示的 T_2 加权像（T_{2W}）。这种成像的典型设置是 TE=100ms，TR=3000ms。如果将长 TR 和短 TE 相结合，那么，就可以产生所谓的质子密度加权像（PD）（见图 16.50c）。其典型的取值是 TE=30ms，TR=3000ms。前面提到过，T_1 对物体不同区域的热特性比较敏感，而 T_2 则对截面上不同区域的磁场非均匀性比较敏感。MRI 检测的定时和脉冲时序设置方法还有很多种，它们在各种不同临床成像应用中都具有重要的作用。

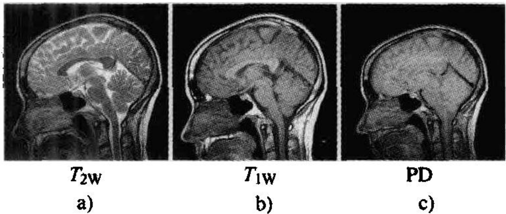


图 16.50 颅脑矢状面的 3 种 MRI 权重图像（图像由 Philips 公司赠送）
a）T_2 加权像 b）T_1 加权像 c）质子密度加权像

对比度是 MRI 系统的一个重要参数，它反映了不同组织之间信号强度的差异。图 16.51 显示了两种不同组织纵向磁化矢量随时间变化的函数，可见两条曲线的上升过程有所不同，表明了两者 T_1 的差别。注意，两条曲线在不同时间点上的垂直间距是不同的，因此，两种组织之间的对比度是随时间变化的。如果比较 400ms、800ms 和 1600ms 这 3 个时间点上两条曲线之间的间距，就可以看出其变化。由此可见，通过改变检测的时间点就可以改变对比度。还有一种提高 MRI 图像对比度的方法就是注射顺磁性物质，如含钆造影剂等。

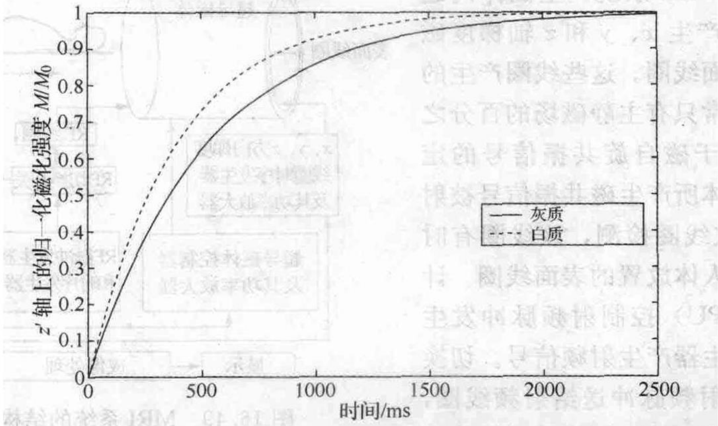


图 16.51 脑组织中灰质和白质的纵向磁化强度随时间变化的曲线

16.3.8 功能磁共振成像

功能磁共振成像（即 fMRI）就是利用 MRI 检测大脑不同区域的活动变化，通常包括大脑的新陈代谢、血流、容积和血氧浓度等的变化，用于考察大脑执行各种不同任务时的响应。这些测量参数的变化都相互关联在一起，并且可能产生相反的作用效果。例如，血流的增加会增加血氧浓度，而新陈代谢的增加却会降低血氧浓度。最常用的检测方法是测量血红蛋白磁化率的变化。氧合血红蛋白是逆磁性物质，而还原血红蛋白是顺磁性物质。利用这两者之间的差别所建立的检测方法被称为血氧水平依赖性（Blood Oxygen Level Dependent，BOLD）磁共振成像，即 BOLD-fMRI。血氧浓度的变化会引起 T_2^* 衰减常数的变化，因此，BOLD-fMRI 采用的是 T_2 加权像。由于这种变化非常微弱，因此，检测时，需要采集大脑执行任务前后的图像，然后将两张图像相减，再将相减的结果叠加在标准图像上。采集图像时还要特别注意去除各种噪声的干扰。一般需要在同一个成像断层上采集数百张图像，然后经过统计处理，才能获得最终的图像。这种检测可能的噪声来源有：热噪声、头部运动以及呼吸周期和心动周期的影响等。图 16.52 显示的就是 BOLD-fMRI 图像。

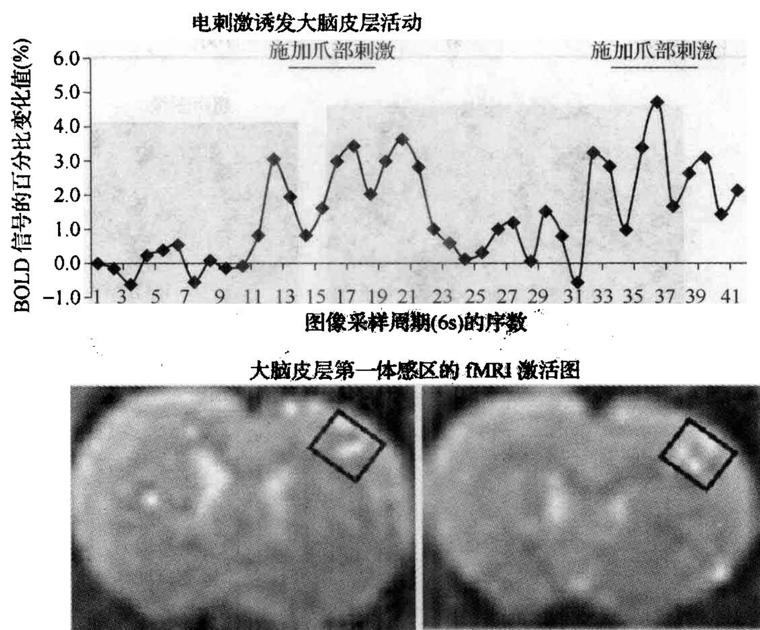


图 16.52 利用快速自旋回波信号的功能磁共振成像

（经 Elsevier 许可，摘自 Ludwig 等，2004。）

注：图中显示的是大鼠左后肢爪部受到电刺激时右侧大脑皮层体感区（激活图方框所示区域）的响应数据，也就是 BOLD 信号的百分比变化值。

16.4 几种成像技术的比较

下面总结主要医学成像技术的特点。表 16.4 列出了这些成像技术的特性。图 16.53 则显示了人体右肾的 MRI、CT 和超声图像。

表 16.4 几种主要医学成像技术之间的比较

	超　声	X 射线	CT	MRI
检测量	力学特性	组织的平均吸收系数	组织的吸收系数	生化参量（T_1、T_2、PD）
观察窗	小窗口即可	人体两侧面	环绕人体一周	环绕人体一周
空间分辨率	0.3～3mm，与频率和方向有关	约 1mm	约 1mm	约 1mm
穿透性	3～25cm，与频率有关	极好	极好	极好
安全性	很好	电离辐射	电离辐射	很好
速度	100 帧/s	数分钟	半分钟～数分钟	数分钟
费用	$	$	$ $ $ $	$ $ $ $ $ $ $ $
便携性	极好	好	不好	不好

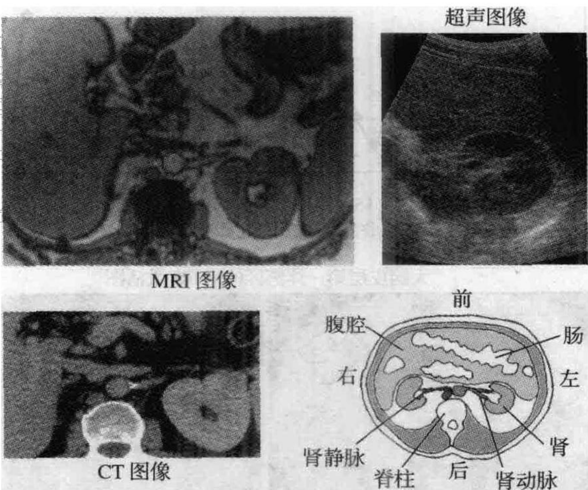


图 16.53 不同成像技术所呈现的右肾图像（由 Marilyn Roubidoux 馈赠）

注：左上是 MRI 图像；左下是 CT 图像；右上是超声图像；右下是显示肾脏位置的腹部横截面示意图。

16.4.1 超声成像技术

与其他成像技术不同，超声图像的分辨率和穿透性取决于超声波的中心频率和

换能器的类型。分辨率随空间位置变化，并且取决于激活孔径的大小、换能器的中心频率和带宽以及声束的聚焦深度。常用的聚焦深度与发射孔径之比是 5，这时，中心频率处声束的半功率宽度约等于两个波长；因此，侧向空间分辨率约为两个波长。对于常用的 1～15MHz 超声波频率范围，侧向分辨率就是 0.3mm～3mm。而且，聚焦区域的分辨率最高。由于尺寸为几个到几十个波长的发射孔径会引起衍射效应，其他区域的分辨率会产生非均匀变化。对于短脉冲，轴向分辨率约为两个波长。

影响超声图像分辨率的另一个主要因素是衰减作用，它限制了声波的穿透深度。中心频率越高，检测深度越深，那么，声波的衰减就越大；因此，在较深的检测区域，声波较弱，难以达到较高的分辨率。

超声图像非常详细，它根据生物组织的“声学特性”，也就是特征阻抗的差异，呈现人体力学结构的图像。特征阻抗取决于组织的硬度、弹性和密度。由于其成像速度可以达到每秒数百帧，超声成像还可以用于显示心脏等人体器官的动态活动。

医学超声是一项非侵入式的检测技术，并且很安全，没有任何生物副作用的积累。超声成像还有两个优点就是费用低且轻便。如今，用于筛查的小型便携式超声仪正在不断推广，超声成像的这两个优点也越来越受到关注。

不过，高质量超声图像需要高水平的操作技术才能获得。这是因为：超声探头的观察窗有众多的选择，人体的解剖结构有差异，观察面也有很多选择，寻找与诊断密切相关的观察面和目标需要一定的经验。另外，图像上生物组织的识别、解释和测量也都需要大量经验的积累。

16.4.2 CT 成像技术

本书第 15 章介绍过 X 射线计算机断层（CT）成像技术，它有时也被称为计算机轴向断层（CAT）成像。X 射线穿过人体时会被组织吸收，从而在射线路径上产生一个总“平均衰减”量。CT 的空间分辨率与射线的波长无关，而是由 X 射线管的焦点大小和生物组织的散射决定，目前的最新技术可达到 1mm。吞服或者注射放射性造影剂可以提高 CT 图像的血管分辨能力。虽然曝光时间很短，但 X 射线毕竟是一种电离辐射，它具有累积剂量效应，对于眼睛等敏感器官或者孕妇都需格外谨慎。

CT 设备庞大且是固定的，可以容纳一个人体，因此，其运行费用相当高。通过 ECG 信号的同步控制，CT 现在已经可用于采集心脏运动的连续图像。CT 扫描机产生的人体大脑、骨骼、肺和软组织的图像质量非常高，可以与超声图像互相补充。

虽然 CT 扫描机的操作也需要经过一定的训练，但并不难。不过，要正确解释 CT 断层图像，并作出明确的诊断，还是需要大量的经验积累。

16.4.3 磁共振成像技术

磁共振成像时，病人躺在巨型封闭电磁体产生的强静磁场中。其图像的分辨率主要取决于磁场梯度，一般为 1mm。成像系统检测的是与磁共振弛豫时间成正比的

感应电压信号，然后应用图像重建算法，根据这些测量数据，计算出人体截面的断层图像。成像速度较快，且比较安全，没有电离辐射。不过，应用时，不要让铁磁性物质接近设备所使用的强大磁体，并且，要知道设备所能提供的磁场强度是有限的，各种磁场的切换速度也是有限的。由于成像设备很贵，因此，MRI 检查费用也很高。

MRI 设备有好几个参数可供选择，如定时、取向和磁场频率等；因此，需要具备高水平的操作技术，才能得到具有诊断意义的图像。并且，只有全面了解成像设备的各种设置，拥有大量的经验，才能正确解释 MRI 图像。

16.4.4 总结

表 16.4 总结了上述 3 种成像技术各自的不同特点。不同成像技术所检测的组织特性完全不同，可以互相补充。CT 和 MRI 都可以用同样的分辨率和对比度对整个人体成像，不过，这两种技术的费用最高，成像速度很慢，设备不能搬动。超声图像的分辨率不一致，穿透深度有限，只能用于人体某些部位的成像，并且，只能对软组织成像，但是它具有低成本、可携带、成像速度快等优点。

习 题

1. 给定傅里叶变换对$\Im[\Pi(\tau)]=\sin(\pi\varphi)/(\pi\varphi)=\mathrm{sinc}(\varphi)$，请求出 $\Pi[(t-t_0)c/L]$ 的傅里叶变换。

2. 下列哪个事件的发生促使医学超声成像开始迅速发展起来？（a） Wild 和 Reid 证明了实时成像的可能性；（b） Edler 和 Hertz 开始研究超声心动图；（c） Stewart 发现儿童患癌症与出生前受到 X 射线照射有关；（d） Satomura 和同事们用超声多普勒技术检测血流。

3. 请利用表 16.1 的数据计算血液与肌肉交界面上垂直入射声波的压力反射系数（单位为 dB），并将其与 45°入射时的反射系数进行比较，哪个数值较大？

4. 请推导垂直入射时声强（而不是声压）的反射系数和透射系数，并分别计算血液与肝、理想刚性边界和空气之间的垂直入射声强透射系数。将这些声强透射系数与相应的声压透射系数进行比较。

5. 假设换能器等效电路的 Q 值定义为 $Q=f_0/\Delta f=1/[\omega_0 C_0 R_A(f_0)]$，其中 Δf 为－6dB带宽，f_0 为 5MHz。请比较以下两种换能器材料的－6dB 带宽百分比数值（即 $\Delta f/f_0$）：（a） $k_T=0.7$，$\varepsilon_R=1470$；（b） $k_T=0.9$，$\varepsilon_R=680$。提示：请参考式（16.27c）。

6. 请利用例 16.3 的常数计算以下两种换能器设计在谐振时的电阻抗：一种是压电晶体直接接触水负载，另一种是在晶体与水之间加匹配层，其中 $Z_B=Z_C$。（提示：利用式（16.34)。）加了匹配层之后，声损耗降低了多少？

7. 超声不能直接用于骨骼成像。（a） 请计算垂直入射时肌肉与骨骼界面上的声

压反射系数；(b) 如果在肌肉与骨骼之间加一层匹配层，请计算匹配层的阻抗；(c) 加了匹配层之后，谐振时的声压反射系数和透射系数各是多少？

8. 某公司只能为其生产的换能器产品提供可调电感或者匹配层，你认为采用哪种方式比较好，为什么？有关数据如下：换能器晶体的尺寸为面积 $A=400\text{mm}^2$，厚度 $d=0.87\text{mm}$；晶体数据为 $c=4.35\text{km/s}$，$\varepsilon_R{}^S/\varepsilon_0=830$，$\varepsilon_0=8.85pF$，$k_T=0.49$；电源阻抗 $R_g=50\Omega$。

9. 请计算以下两种超声扫描图像的帧频：(a) 扫描深度为 10cm、每帧 128 线的 3MHz 心脏扇形扫描仪；(b) 扫描深度为 1cm、每帧 300 线的 10MHz 线阵扫描仪。

10. 假设某成像系统 5MHz 换能器的总的往返动态范围为 100dB。如果用它检测肝脏与骨骼界面的声压反射，那么，最大检测距离为多少？对于肝脏与肌肉界面，最大检测距离又是多少？假设界面之前的平均组织特性与肝脏相同。有关组织的特性参数请参考表 16.1。

11. 假设需要探查 6cm 深的一块直径约为 2mm 的损伤区，有下列几种具有不同孔径和中心频率的线阵换能器可供选用，你会选择哪一种，为什么？(a) $L=15\text{mm}$，$f_c=6\text{MHz}$；(b) $L=11\text{mm}$，$f_c=5\text{MHz}$；(c) $L=12\text{mm}$，$f_c=3\text{MHz}$。假设 $c=1.5\text{mm}/\mu\text{s}$，扫描平面内的聚焦长度 $F=60\text{mm}$。

12. 对于圆形孔径为 $a=5\text{mm}$，频率为 5MHz 的换能器，配上以下哪种几何聚焦长度的透镜之后可以使其总的有效聚焦长度达到 40mm？(a) $F=50\text{mm}$；(b) $F=40\text{mm}$；(c) $F=80\text{mm}$。

13. 某 2MHz 血液多普勒超声系统的检测阈值为 −35dB。假设血管壁外附着的肌肉层吸收损失可以忽略不计，血流速度为 1m/s。请计算可以提供足够信号强度的入射角范围（精确到°）以及相应的多普勒频率。有关数据请参考表 16.1。注意，多普勒角度＝90°－入射角。

14. 检测系统如图 16.29 所示。测量旋转磁体所产生的感应电压时，如果增加线圈的半径 d，那么，所测得的电压会增大还是减小？请解释其物理原理。注意，电压单位 V＝Wb/s。

15. 请利用表 16.2 的数据分别计算 ^{1}H、^{19}F 和 ^{23}Na 的所有核磁偶极矩 μ_z。

16. 设温度 $T=300\text{K}$，磁场 $B_0=0.5\text{T}$，且质子密度 $N=6.7\times10^{19}$ 质子/mm^3；请计算水的净磁化矢量 $\boldsymbol{M}_0$。利用式 (16.5) $\hat{\boldsymbol{M}}_0=\sum_n \hat{\boldsymbol{\mu}}_n=\hat{\boldsymbol{\mu}}(n_+-n_-)$，并假设 $n_+=N/2$。

17. 假设 $\hat{\boldsymbol{B}}_y(t)=\hat{\boldsymbol{y}}B_1\sin(\omega t)$，那么，旋转坐标系中的磁场将是什么？低通滤波除去 2ω 频率之后，剩下的各个分量是什么？

18. (a) 设 $B_1=10^{-5}\text{T}$，如果要产生 $\pi/2$ 弧度的翻转角，那么，^{1}H 所需的脉宽（即翻转时间）t_p 是多少？(b) $\pi/2$ 的翻转角再接一个 π 旋转，请计算其总的翻转时间。

19. 请分别计算 $t=250\text{ms}$ 时刻大脑白质和恶性肿瘤的 $\pi/2$ 和 π 翻转角所产生的

相对弛豫。哪个翻转角下两种组织之间的对比度较大？再比较 $t=500\text{ms}$ 时刻的情况。

20. (a) 请推导 FID 脉冲 $v'(t)=v(t)H(t)$ 的频谱表达式。其中 $H(t)$ 是阶跃函数，也就是，$t>0$ 时，$H(t)=1$；$t<0$ 时，$H(t)=0$。利用傅里叶变换对$[e^{-a|t|}H(t)]=\dfrac{1-i2\pi f/a}{|a|[1+(2\pi f/a)^2]}$。(b) FID 频谱与自旋回波的复数比是什么？(c) 请计算该复数比在拉莫尔频率处的值。(d) 请根据两者的时间信号波形解释这个比值。

21. (a) 如果习题 20 所述 FID 脉冲的 $T_2=T_2^*=5\text{ms}$，请计算 5MHz 拉莫尔频率处的频谱分量。设 $B_1=20\mu\text{T}$，$I=1\text{A}$，$V_{\text{voxel}}=1\text{cm}^3$，$M_{xy}^0=5\times10^{-12}\text{J}/(\text{T}\cdot\text{mm}^3)$。(b) 对于脂肪组织自旋回波的 T_2，在相同拉莫尔频率处的频谱分量又是多少？

22. 假设需要扫描 z 轴上中心位置前后 $-15\text{cm}\sim15\text{cm}$ 范围内的 1mm 层厚的 MRI 断层图像，请计算所需时间脉冲包络的 -6dB 带宽，以及带宽两端的频率。设 $G_z=0.5\text{T/m}$。

23. 为了覆盖 y 轴上中心位置前后 $\pm10\text{cm}$ 的扫描范围。对于 ^1H 用 $t_p=10\mu\text{s}$ 的脉冲进行相位编码，总相移为 $\pi/2$。请计算磁场两端的梯度系数 G_{ym}，以及 256 线分辨率时的梯度系数变化率。

24. 给定 $f_n=78.5\text{MHz}$，$G_x=10^{-4}\text{T/m}$，$t_p=5\mu\text{s}$，$u_m=21.3\text{m/T}$；对于 ^1H，请计算测量点 (x_n,y_m) 相对于中心点的位置。

25. 请根据式 (16.87) 推导 $s_{n,m}(k_x, k_y)$。

26. 请回答下列有关 MRI 成像的问题：

a. MRI 图像显示的是什么信号？

b. 如何根据一系列测量信号确定人体内部空间位置？

c. 在三个成像轴中，哪些是可以主动控制的，哪些是被动检测的？

27. 请解释 T_1、T_2 和 PD 加权像区分各种生物组织的基本物理原理。

推荐阅读资料

Bitter, F. and Medicus, H.A. (1973). *Fields and Particles: An Introduction to Electromagnetic Wave Phenomena and Quantum Physics*. American Elsevier, New York.

Chakeres, D.W. and Schmalbrock, P. (1992). *Fundamentals of Magnetic Resonance Imaging*. Williams & Wilkins, Baltimore, MD.

Cho, Z.H., Jones, J.P. and Singh, M. (1993). *Foundations of Medical Imaging*. Wiley, New York.

Jensen, J.A. (1996). *Estimation of Blood Velocities Using Ultrasound*. Cambridge Univ. Press, Cambridge.

Kim, E.E. and Jackson, E.F. (1999). *Molecular Imaging in Oncology*. Springer-Verlag, Berlin.

Kino, G.S. (1987). *Acoustic Waves: Devices, Imaging, and Analog Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ.

Kuperman, V. (2000). *Magnetic Resonance Imaging Physical Principles and Applications*. Elsevier Academic, Boston.

Lauterbur, P.C. (1973). Image formation by induced local interactions: Examples employing

nuclear magnetic resonance. *Nature* 242, 190–191.

Liang, Z.P. and Lauterbur, P.C. (2000). *Principles of Magnetic Resonance Imaging: A Signal Processing Approach*. IEEE Press, New York.

Ludwig, R., Bodgdanov, G., King, J., Allard, A. and Ferris, C.F. (2004). A dual rf resonator system for high-field functional magnetic resonance imaging of small animals. *J. Neurosci. Meth.* 132, 125–135.

Panda, R.K. (1998). Development of novel piezoelectric composites by solid freeform fabrication techniques. Dissertation. Rutgers, New Brunswick, NJ.

Shung, K.K., Smith, M.B. and Tsui, B.M.W. (1992). *Principles of Medical Imaging*. Academic, San Diego, CA.

Szabo, T.L. (1998). Transducer arrays for medical ultrasound imaging. In *Ultrasound in Medicine, Medical Science Series, F* (A. Duck, A.C. Baker and H.C. Starritt, Eds.). Institute of Physics Pub., Bristol, U.K.

Szabo, T.L. (2004). *Diagnostic Ultrasound Imaging: Inside Out*, Elsevier Academic, Boston.

第 17 章　光学和激光在生物医学中的应用

Gerard L. Cote 博士

LiHong V. Wang 博士

Sohi Rastegar 博士

本章目录

本章主要内容：

- 基本光学原理、光在生物组织以及其他生物和生化介质中传播的基本理论；
- 根据基本工程原理开发的激光等光学技术的检测、成像、诊断和治疗仪器设备；
- 光纤等光学系统与生物组织之间的生物化学和生物物理作用；
- 激光与生物组织之间作用的光热效应。

从早期利用日光进行简单的热疗以及眼睛、皮肤和伤口的检查，到如今利用激光和内窥镜等先进技术实现各种医学诊断和治疗，光在医学中的应用一直在不断发展。的确，近 10 年，光学技术给临床医学带来了许多变革。与其他技术相比，光学技术的低成本使它在医疗领域的应用具有巨大的发展潜力。例如，应用光纤传像术的腹腔镜胆囊切除术，是一种创伤最小的胆囊切除手术，它将原来痛苦且昂贵的手术几乎变成了一种门诊手术。经过不断的努力，生物医学工程师们已经掌握了光和热与生物介质之间相互作用的许多基本原理，光学在皮肤学、眼科学、心脏病学和泌尿学等医学领域的应用正在迅速发展。

传统光学工程是电气工程课程的一部分，它主要讲述光学在国防技术中的应用。将光学应用于生物医学时，会涉及到很多光与生物组织之间的相互作用问题，这些问题在传统光学教材和课程中并不介绍。因此，本章的目的是较详细地给学生们讲解日益发展的生物医学光学的基本理论，以及基于光学技术的各种监测、诊断和治疗仪器。

17.1　基本光学理论概述

自古以来，伽利略、牛顿、惠更斯、菲涅尔、麦克斯韦、普朗克和爱因斯坦等许多伟人都研究过光的性质。开始时，人们认为光的本质不是粒子就是波。但是，经过长期的争论和研究之后，人们得出的结论是：在某些情况下，光表现出波动性；而在另一些情况下，光又会表现出粒子性。本节将从波动方程开始，阐述光的基本特性，并且介绍光谱以及与光的吸收、散射和偏振相关的控制方程。其中，波动方程一般由麦克斯韦方程推导得到。

17.1.1 电磁波和光谱

19世纪末，麦克斯韦（J. Clerk Maxwell）创立了电磁学的几个基本定理，并由此推导出波动方程，得出光的本质是电磁波的结论。此推论的关键是波动方程真空解所对应的电磁波的速度正好与实验测得的光速相等。很多有关光学的导论性教材（如 Hecht 和 Pedrotti 的著作等）都是从以下矢量形式的麦克斯韦波动方程开始讲述：

$$\nabla^2 E=\varepsilon_0\mu_0\partial^2 E/\partial t^2 \tag{17.1}$$

式中 $\nabla^2 E$——电场强度对于空间位置的二阶偏导数，即 $\partial^2 E/\mathrm{d}x^2+\partial^2 E/\mathrm{d}y^2+\partial^2 E/\mathrm{d}z^2$；

ε——介质的介电常数；

μ——介质的磁导率。

在真空中，$\mu=\mu_0$，且 $\varepsilon=\varepsilon_0$。

由对称性可知，磁场强度 $\boldsymbol{H}$ 也具有同样形式的方程。

波速 c_0 等于光速，其值已知，可用如下表达式表示：

$$c_0=1/(\varepsilon_0\mu_0)^{1/2} \tag{17.2}$$

因此，麦克斯韦波动方程描述了电场强度（或磁场强度）的空间变化与速度以及场强的时间变化之间的关系。在图17.1所示的笛卡尔坐标 O_x、O_y、O_z 下，此波动方程的一个特解就是以下正弦波函数：

$$E_x=E_0\exp[\mathrm{j}(\omega t-kz)] \tag{17.3}$$

$$H_y=H_0\exp[\mathrm{j}(\omega t-kz)] \tag{17.4}$$

这两个方程表明电场是 x-z 平面上的正弦波，而磁场则是与电场正交的 y-z 平面上的正弦波，两个波均沿 O_z 方向传播。

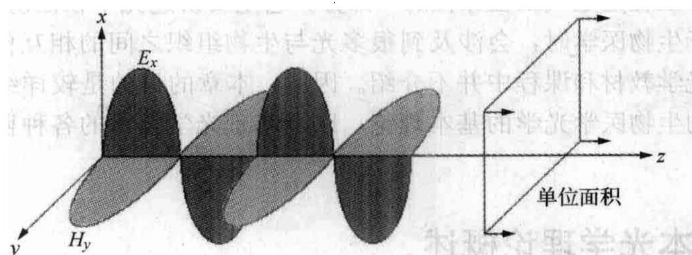


图 17.1 正弦电磁波

电磁波的频率 f 和波长 λ 都与速度 c 有关，即

$$f=\omega/2\mathrm{p} \tag{17.5}$$

$$\lambda=2\mathrm{p}/\mathrm{k} \tag{17.6}$$

$$c=f\lambda=\omega/\mathrm{k} \tag{17.7}$$

电磁波在真空中的传播速度 c_0 与在介质中的传播速度 c 之比被称为介质的折射率 n，

即 $n=c_0/c$。

例 17.1　红色氦氖激光的波长为 633nm，请计算其在折射率为 1.33 的清澈角膜组织中的传播速度。在折射率为 1.5 的玻璃中，该激光的传播速度又会变成多少？并请说明该计算结果的意义。

解：由 $n=c_0/c$ 可得

对于角膜组织：$c_{组织}=c_0/n_{组织}\approx(2.998\times10^8)\text{m/s}/1.33=2.25\times10^8\text{m/s}$

而对于玻璃则：$c_{玻璃}=c_0/n_{玻璃}\approx(2.998\times10^8)\text{m/s}/1.5=2.00\times10^8\text{m/s}$

该结果说明：光在清澈的生物组织等具有较小折射率的介质中的传播速度比玻璃中要快。本章后面将计算光线穿过各种不同介质，如眼球的角膜、房水和晶状体等清澈组织时，光线的反射角和折射角，那时就可以进一步理解本例题计算结果的意义。■

电磁波的辐射强度，也就是单位面积的功率，可以用坡印亭（Poynting）矢量 $\boldsymbol{\Pi}$ 来表示，它定义为电场与磁场的矢量乘积：

$$\boldsymbol{\Pi}=\boldsymbol{E}\times\boldsymbol{H} \tag{17.8}$$

电磁波一个周期上坡印亭矢量的平均值就是电磁波的辐射强度 I，因此，如图 17.1 所示，如果 E 和 H 在空间上是正交的，那么

$$I=\langle|\Pi|\rangle=\text{单位面积功率}=c\varepsilon E^2 \tag{17.9}$$

式中　$\langle|\Pi|\rangle$——$\boldsymbol{\Pi}$ 的平均值。

由此可见，电磁波的辐射强度与电场强度的次方成正比，这在各向同性的绝缘介质中成立。由麦克斯韦方程可知，辐射强度同样也与磁场二次方成正比。因此，如果电场与磁场在相位上正交，即相位差为 90°，那么，由于正弦与余弦乘积的平均值为零，因此，电磁波的辐射强度将等于 0，即

$$I=\langle|\Pi|\rangle=\langle E_0\cos(\omega t)H_0\sin(\omega t)\rangle=0 \tag{17.10}$$

除了辐射强度以外，光还可以用波长特性来描述，也就是电磁波谱，其频率范围从 10^3 Hz 的低频无线电波到 10^{20} Hz 的 γ 射线。如图 17.2 所示，光的波长范围可以分成以下几个区段：波长为 0.003～0.4μm 的紫外线，波长为 0.4～0.7μm 的可见光（其相应的频率范围为 $7.5\times10^{14}\sim4.3\times10^{14}$ Hz），波长为 0.7～2.5μm 的近红外光，波长为 2.5～12μm 的中红外光，以及波长大于 12μm 的远红外光。

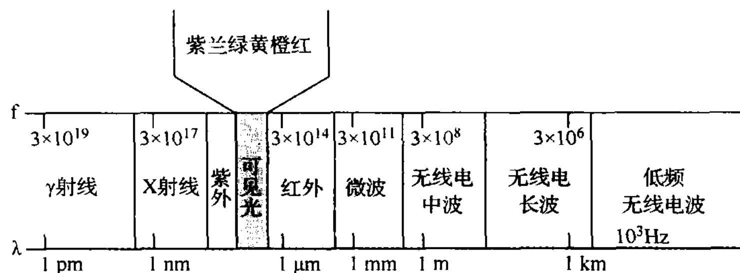


图 17.2　电磁波的波谱

17.1.2 光的偏振

上节所述的 E_x 和 H_y 正弦函数解只是麦克斯韦方程无穷多个正弦解之一。角频率为 ω 的正弦通解可以写为

$$\boldsymbol{E}(\boldsymbol{r},\ t)=\boldsymbol{E}(\boldsymbol{r})\exp(\mathrm{j}\omega t) \tag{17.11}$$

式中 $\boldsymbol{E}(\boldsymbol{r},\ t)$，$\boldsymbol{E}(\boldsymbol{r})$——复数矢量；

$\boldsymbol{r}$——实数半径矢量。

为了简化，这里只考虑真空中沿 O_z 方向传播的平面波单色光，即单波长光，其电场波动方程的通解可以写为

$$E_x=e_x\cos(\omega t-kz+\delta_x) \tag{17.12}$$

$$E_y=e_y\cos(\omega t-kz+\delta_y) \tag{17.13}$$

式中 δ_x 和 δ_y——任意相位角。

这样，此解用 x-y 平面和 y-z 平面上的两个波就可以完整地描述。在某个特定 z 值点 z_0 处，这两个波的方程分别是：

$$E_x=e_x\cos(\omega t+\omega\delta'_x)\ \delta'_x=\delta_x-kz_0 \tag{17.14}$$

$$E_y=e_y\cos(\omega t+\delta'_y)\ \delta'_y=\delta_y-kz_0 \tag{17.15}$$

这就好像每个电场矢量的端点在沿着一条直线随时间做正弦振动。其中，E_x 被称为 O_x 方向的线性偏振，而 E_y 被称为 O_y 方向的线性偏振。

由式（17.15）可见，对于完全偏振光，只要用幅值和相位，即 2×1 矩阵 $[e_x\exp(\mathrm{j}\delta'_x)\ e_y\exp(\mathrm{j}\delta'_y)]$，就可以完整地描述其方程。这种矢量表示法被称为 Jones 矢量。光学系统中，如果光的传播介质是无散射（因而不存在去偏振作用）的透镜或者清澈的生物组织等，那么，这种介质就可以用 2×2Jones 矩阵完整地描述出来。而偏振光穿过这种介质之后的输出光就可以用 Jones 矩阵与输入光矢量的乘积来表示。如果光的传播介质具有散射性，也就是传播的光既包含全偏振光又包含部分偏振光，那么，这种系统可以用 4×4 的 Mueller 矩阵和 4×1 的 Stokes 矢量来描述。

再来看式（17.15），偏振光矢量端点的轨迹是 E_x 和 E_y 两个矢量之和，一般，它具有图 17.3 所示的椭圆运动轨迹，在 $z=z_0$ 处的 x-y 平面上，其笛卡尔坐标方程为

$$E_x^2/c_x^2+E_y^2/c_y^2+2(E_xE_y/e_xe_y)\cos\delta=\sin^2\delta \tag{17.16}$$

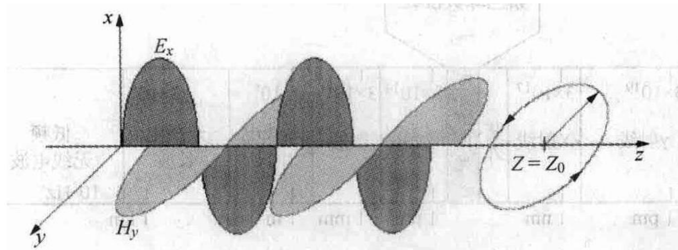


图 17.3 电磁波的偏振传播

其中，$\delta=\delta'_y-\delta'_x$。应该注意，如果（1）$E_x\neq0$ 且 $E_y=0$；或者（2）$E_y\neq0$ 且 $E_x=0$；或者（3）$\delta=m\pi$（m 为正整数或负整数），那么，椭圆运动轨迹就会变成直线，从而产生线偏振光。如果 $e_x=e_y$ 且 $\delta=(2m+1)\pi/2$，由于 E_x 和 E_y 具有相同的幅值且相位正交，椭圆轨迹就变成了圆。

对于物理性质与方向相关的各向异性介质，光的偏振特性尤其重要，此时，E_x 与 E_y 不相等，光传播路径上 e_x、e_y 和 δ 都有变化，双折射晶体或保偏光纤就属于这类各向异性介质。

例 17.2 设波动方程的一个通解是如下表达式所表示的平面电磁波：

$$E_x=0,E_y=0,E_y=2\cos[2\pi\times10^{14}(t-x/c)+\pi/2]$$

请计算电磁波的频率 f、波长 λ、运动方向（即 x、y 或 z）、幅值 A、初始相位角 ϕ 以及波的偏振性（即线偏振、圆偏振或者椭圆偏振）。注：使用 SI 国际单位制。

解： 平面电磁波通解可以表示为 $E=A\cos(kx\pm\omega t+\phi)$，其中 $k=2\pi/\lambda$，$\lambda=c/f$，$\omega=2\pi f$，$c=3\times10^8\text{m/s}$。将题中给出的通解表达式变换为 $E_y=2\cos[2\pi\times10^{14}t-(2\pi\times10^{14}x/c)+\pi/2]$。对比这两个方程可知：$f=1\times10^{14}\text{Hz}$，$\lambda=3\times10^{-6}=3\mu\text{m}$，且电磁波沿 x 轴正向传播，$A=2\text{V/m}$，$\phi=\pi/2$。由于 $E_x=E_z=0$，此电磁波必定是 y 方向的线偏振波。■

17.1.3 光的吸收、散射和发射

讲述了光的偏振性质之后，下面重点介绍用于描述光穿过生物介质时所发生变化的光学特性，包括光的吸收、散射和各向异性。这些光学特性在光学诊断、传感器开发以及光学治疗方面都有许多应用。例如，有人正在利用光的吸收特性开发非侵入式定量检测人体葡萄糖等化学物质的方法；另外，还有人在利用光的吸收和散射效应，模拟前列腺等手术过程中特定波长的光所产生的凝血和组织消融作用。

光在生物组织中的作用深度和作用体积是光的吸收和散射特性的函数。如果光的散射可以忽略不计，也就是以光的吸收为主，那么，透射光强度的变化由光路的吸收特性和入射光的波长决定。由于生物组织的含水量很高，波长大于 1.5 μm 的光穿过各种生物组织时都是以光的吸收为主。光强与光程和溶液浓度之间的关系可以用如下比尔-朗伯定律来描述：

$$I(z)=I_0\exp(-\mu_a z) \tag{17.17}$$

式中 I——透射光强度；

I_0——入射光强度；

z——光程长度；

μ_a——吸收系数，单位为 1/cm。

吸收系数可以进一步分解为吸光物质的浓度 C（mg/mL）与其摩尔吸收系数 v（cm^2/mg）的乘积，而 v 则是入射光波长的函数。

就物体的冷发光而言，根据发光寿命的长短，可分为荧光和磷光。当电磁辐射的光子被物体分子吸收之后，这些分子会跃迁到某个高能级激发态，随后再返回到

低能级，在此过程中就会发射出辐射光，这就是物体的发光过程。荧光的发射不涉及电子自旋状态的变化，因此产生速度较快，寿命较短。物体发光过程中所吸收的能量可以用单光子能量来描述，其公式为

$$E=hc/\lambda \tag{17.18}$$

式中 E——单光子能量；

h——普朗克常数，其值为 6.626×10^{-34} J·s；

c——光速；

λ——波长。

由于光子发射之前原吸收能量中的一部分已被耗散，因此，这种吸收和重发射过程并不能达到100%的效率，激发光的能量总是比发射光的能量大，即

$$E_{激发}>E_{发射} \tag{17.19}$$

并且，发射光的波长总是大于激发光的波长，即

$$\lambda_{发射}>\lambda_{激发} \tag{17.20}$$

荧光物质的发射光能级由其分子发射光子的离散能级决定，每种特定的荧光染料在某个给定波长激发光的作用下，只能发射某个特定波长的荧光。荧光辐射原本只是用于检测稀溶液的浓度，是一种灵敏度极高且成本很低的技术。最近，研究人员开始将荧光辐射用于生物组织等混浊光学介质的检测和诊断。

除了上述光的吸收和发射之外，生物组织还具有光的散射作用，尤其对于可见光和近红外波长的光。与光的吸收和发射不同，光的散射不涉及原子或者分子在量子化能级之间的能量跃迁，它只是由介质介电常数在空间的随机变化产生的作用。如果存在光的散射效应，实际的光强分布就会与比尔定律的计算结果截然不同。光线穿过生物组织时，会经历大量散射事件，使光从平行光束中发散出去。已经证明，平行光束穿过既有吸收又有散射的混浊介质时所产生的能量转换可以用光的传输理论来描述。该传输理论以前一直用于描述光的散射、吸收和荧光辐射，近来也用于描述光的偏振特性，其原因之一是只需经历没几次散射事件，光束的偏振就可以被完全随机化。17.2.2节将介绍这种传输理论，它是基于多重散射介质中光子传输过程的统计近似，是一种直观推断方法。

17.2 生物组织中光传播的基本理论

本节将介绍光在生物组织中的传播特性。首先不考虑介质的吸收和散射效应，讲述光在介质中的几何光学传播理论；然后，介绍光传播中的吸收和散射效应，以及边界效应，并讲述各种测量光吸收和散射特性的方法。

17.2.1 光在介质中的几何传播

前面，我们将光定义为沿 O_z 方向传播的一组电磁波。本节我们要把光看做一组直线传播的“光线”，它们由前述复指数描述的平面波构成。“光线”的方法（即几

何光学）对于研究光在两种不同光学介质交界面处的效应非常重要。在几何光学中，假设入射光、反射光和折射光都处于同一平面。如图 17.4 所示，光线到达两种不同介质的交界面时，其传播规律遵循两个基本定理：反射角等于入射角；折射角的正弦值与入射角的正弦值之比为常数（即 Snell 折射定理）。注意，虽然这种光线传播理论可以满足透镜和棱镜等宏观光具设计中的大多数应用情况，但是，它还是有很大的局限性。例如，它不能用于估计折射光和反射光的强度，也不能用于分析光的吸收和散射效应。另外，当光学器件的孔径比光的波长小时，还会产生衍射现象，使光束发散，从而使几何光学理论不成立。

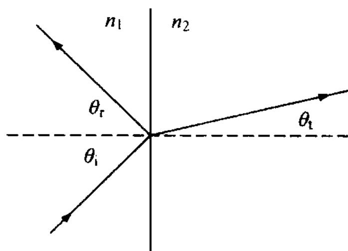
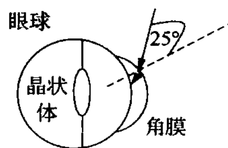


图 17.4　光线在两种不同介质交界面处的传播

注：显然，入射角等于反射角，即 $\theta_i=\theta_r$；由 Snell 折射定理可知，折射角的正弦值与入射角的正弦值之比为常数，即 $n_1\sin(\theta_i)=n_2\sin(\theta_t)$。

例 17.3　如右图所示，假设有一束绿色光线射向眼睛的角膜，光线与界面的法线成 25°角。设空气的折射率为 1.000，角膜的折射率为 1.376。(a) 请计算光线从角膜前表面射向角膜内的出射角；(b) 角膜对于光线有怎样的折射效应？



解：(a) 根据 Snell 折射定理，$n_1\sin(\theta_1)=n_2\sin(\theta_2)$，由此可得角膜内光线的出射角 θ_2 的计算公式为

$$\theta_2=\arcsin(1.000\sin(25)/1.376)=17.89°$$

(b) 由 θ_2 的计算值可知，角膜将光线折向法线。这是合理的，因为眼睛的作用之一就是要使光线产生折射，让光线能够穿过眼球中央的虹膜和晶状体，到达视网膜，从而由大脑形成视觉图像。■

波动光学的理论可以用于计算光线从一种介质传播到另一种介质时反射光和折射光的强度。如图 17.4 所示，假设两种介质都是绝缘体，其边界条件遵循麦克斯韦方程，也就是 $\boldsymbol{E}$ 和 $\boldsymbol{H}$ 的切线分量在边界上连续，即 $\boldsymbol{E}_i+\boldsymbol{E}_p=\boldsymbol{E}_t$，并且磁通量密度 $\boldsymbol{B}$ 和电位移 $\boldsymbol{D}$ 的法线分量也在边界上连续。只有当所有光波的频率都相同时，这些条件才能在任何时刻和边界上任意位置处成立。这样，就可以计算各个光波的相对幅值。注意，$\boldsymbol{E}$、$\boldsymbol{H}$、$\boldsymbol{D}$ 和 $\boldsymbol{B}$ 的值与入射光的电场 $\boldsymbol{E}$ 和磁场 $\boldsymbol{H}$ 相对于入射平面的振动方向有关，也就是与光波的偏振性相关。将入射光分解为两个正交的偏振光，即振动面平行于入射平面的线偏振光（p 分量）和振动面垂直于入射平面的线偏振光（⊥分量）。同时应用麦克斯韦方程、Snell 折射定理和反射定理，再加上边界条件，就可以

推导出以下 Fresnel 方程。其中，平行偏振光的反射方程和透射方程分别为

$$r_p=\left(\frac{E_t}{E_i}\right)_p=(-n_1\cos\theta_t+n_2\cos\theta_i)/(n_1\cos\theta_t+n_2\cos\theta_i) \quad (17.21)$$

$$t_p=\left(\frac{E_t}{E_i}\right)_p=(2n_1\cos\theta_i)/(n_1\cos\theta_t+n_2\cos\theta_i) \quad (17.22)$$

垂直偏振光的反射方程和透射方程分别为

$$r_\perp=\left(\frac{E_r}{E_i}\right)_\perp=(n_1\cos\theta_i-n_2\cos\theta_t)/(n_1\cos\theta_i+n_2\cos\theta_t) \quad (17.23)$$

$$t_\perp=\left(\frac{E_t}{E_i}\right)_\perp=(2n_1\cos\theta_i)/(n_1\cos\theta_i+n_2\cos\theta_t) \quad (17.24)$$

应用这些方程时，要注意如下两种特殊情况的限制。

第一种情况：如果 $n_2>n_1$，则 $\theta_i>\theta_t$，对于任意 θ_i 值，$r_\perp$ 总是负值，而 r_p 则从 $\theta_i=0$处的正值开始逐渐减小，直到 $\theta_i+\theta_t=90°$时，r_p 等于 0。此时意味着折射光和反射光彼此正交，根据 Shell 折射定理，这种情况只有当 $\tan\theta_i=n_2/n_1$ 时才能出现。这个特定的入射角 θ_i 被称为偏振角，也常称为 Brewster 角。在此入射角下，只有垂直于入射平面的偏振光才被反射，而平行于入射平面的偏振光无反射，因此，可以用这种方法产生偏振光。如果 θ_i 增大且超过 θ_p，那么，r_p 就变得越来越负，并在 $\theta_i=90°$时变为-1.0，这意味着交界面在此入射角下变成了全反射镜面。另一方面，如果光波垂直入射，即 $\theta_i=\theta_r=\theta_t=0$，那么 $t_p=t_\perp=2n_1/(n_1+n_2)$，并且 $r_p=-r_\perp=(n_2-n_1)/(n_1+n_2)=(n_1-n_2)/(n_1+n_2)$。

光波的辐射强度可以用检测器实际测得，其大小与电场强度的平方成正比，即 $(I_R/I_i)=(r_p)^2=\left(\frac{E_r}{E_i}\right)^2=(n_2-n_1)^2/(n_1+n_2)^2$，$(I_t/I_i)=(t_p)^2=\left(\frac{E_t}{E_i}\right)^2=4n_1^2/(n_1+n_2)^2$。这两个表达式很有用，它们表示了光从一种介质传向另一种介质时，法向全反射所引起的能量损失。

第二种情况：如果 $n_2<n_1$，则 $\theta_t>\theta_i$，对于任意 θ_i 值，$r_\perp$ 总是正值，且其值从 $\theta_i=0$ 时的初始值开始逐渐增大，直到$+1.0$ 为止，此时的入射角 θ_i 称为临界角 θ_c。当 $\theta_i=\theta_c$ 时，$\theta_t=90°$，由 Shell 折射定理可知，此时 $\sin\theta_t=n_1/n_2\sin\theta_i$。显然，当光线从低密度介质传向高密度介质，即 $n_1>n_2$ 时，由该式计算得到的 $\sin\theta_t$ 值可能大于 1，这对于任何 θ_t 都是不可能的。当 $n_1/n_2\sin\theta_c=1$，也就是 $\theta_t=90°$时，入射角 θ_i 就等于临界角 θ_c。这样，任何 $\theta_i>\theta_c$ 的入射光都会在交界面上产生全反射。这就是光纤的工作原理，光纤的纤芯材料（即光传播的介质）的折射率比其外周包层材料的折射率稍大，因此，光线进入光纤之后，就会在纤芯内不断产生全反射，使传播过程中光的衰减达到最小，光就可以沿着光纤传播出去。

17.2.2 光与介质之间的相互作用——吸收和散射

激光等入射光照射在生物组织上时，生物组织对入射光会产生不同程度的反射、吸收、散射和透射作用。分析这种复杂情况的理想方法是使用能够阐述光分布的电

磁波理论。不过，这种方法的实际应用比较困难，而且还必须已知生物组织各种电特性的可靠数据。还有一个比较实用的方法就是利用传输理论，建立光强传播的微分方程。本节下面将先推导纯吸收介质中光强分布的方程，然后再进一步介绍既具有散射又具有吸收的介质中的光传输的一般方程。

1. 纯吸收介质

为了描述激光能量在不透明吸收性介质中的分布和传输，可以假设介质中同时存在两个“相”：一个是系统的物质相；另一个是电磁辐射的光子相。如图 17.5 所示，物质相用圆圈表示，光子相则用撞击物质相的曲线箭头表示。物质相的能量平衡方程将在 17.3 节和 17.5 节讲述系统的热力学过程时再介绍。这里先讲解光子相的能量平衡方程。

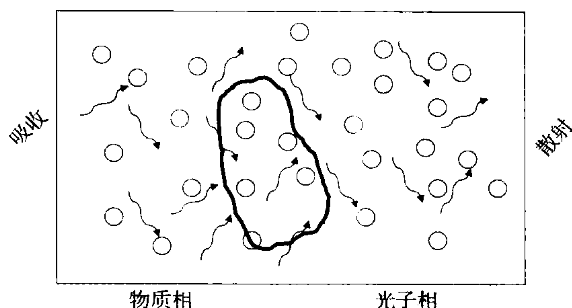


图 17.5　物体中同时存在的两个相：物质相和光子相

如图 17.6 所示，假设某个无穷小的物体单元受到光辐射的照射。辐射能量 $U^{(\mathrm{rad})}$ 随时间的变化率等于射线的入射通量与出射通量之差，再减去物质相的能量吸收率。其中，射线入射通量与出射通量之差的极限就是辐射通量散度的负值。于是，如果用 Q_L 表示物质相的能量吸收率，也就是激光能量转换成热能的部分（即吸收率），用 $\vec{q}^{(\mathrm{rad})}$ 表示辐射通量，那么，光子相的能量变化率就可以用以下微分方程来表示

$$\partial U^{(\mathrm{rad})}/\partial t=-\vec{\nabla}\cdot\vec{q}^{(\mathrm{rad})}-Q_L \tag{17.25}$$

这里忽略了物质相可能存在的光子发射和散射。

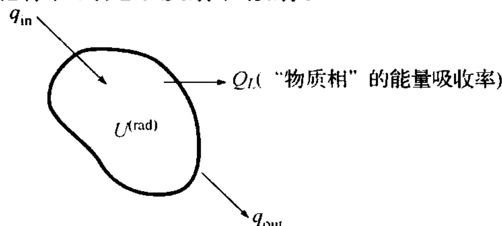


图 17.6　无穷小物体单元中光子相能量变化的示意图

注：图中显示了光辐射的入射通量和出射通量、辐射能量以及由物质相的吸收引起的能量损耗率。

对于单色准直平行入射光的稳态系统（稳态意味着随时间的变化为 0），假设光辐射仅沿 z 轴的正向传播，即只有一个向前的辐射通量，那么，式（17.25）可以简化为

$$dq_z^{(\mathrm{rad})}/dz=-Q_L \tag{17.26}$$

考虑光的传播速度和生物组织的大小，除了时间分辨法以外，对于一般的应用情况，这种稳态系统的假设条件是合理的。这里关键的一步是要利用以下唯象关系式

$$Q_L=\mu_a I \tag{17.27}$$

式中 μ_a——吸收系数；

I——总的光强。

物体上某点的总光强是该点受到的各个辐射通量之和。在此我们只考虑纯吸收组织，只有单个辐射通量，因此

$$I=q_z^{(\mathrm{rad})} \tag{17.28}$$

将式（17.27）和式（17.28）代入式（17.26），可得以下微分方程

$$\frac{dI}{dz}=-\mu_a I \tag{17.29}$$

该微分方程有简化解

$$I(z)=I_0\exp(-\mu_a z) \tag{17.30}$$

其中 I_0 是物体表面 $z=0$ 处的光强。此方程就是本章前面已经提到过的描述纯吸收介质的著名的比尔-朗伯定律，也就是式（17.17）。

由式（17.27）和式（17.30）可知，转换成热能的那部分光能量（即吸收率）现在可以写成

$$Q_L=\mu_a I_0\exp(-\mu_a z) \tag{17.31}$$

此方程可以推广到轴对称的三维空间，用于描述径向对称分布的激光光束垂直入射到平板上的情况，此时方程变为

$$Q_L(r,\ z)=\mu_a I_0\exp(-\mu_a z)f(r) \tag{17.32}$$

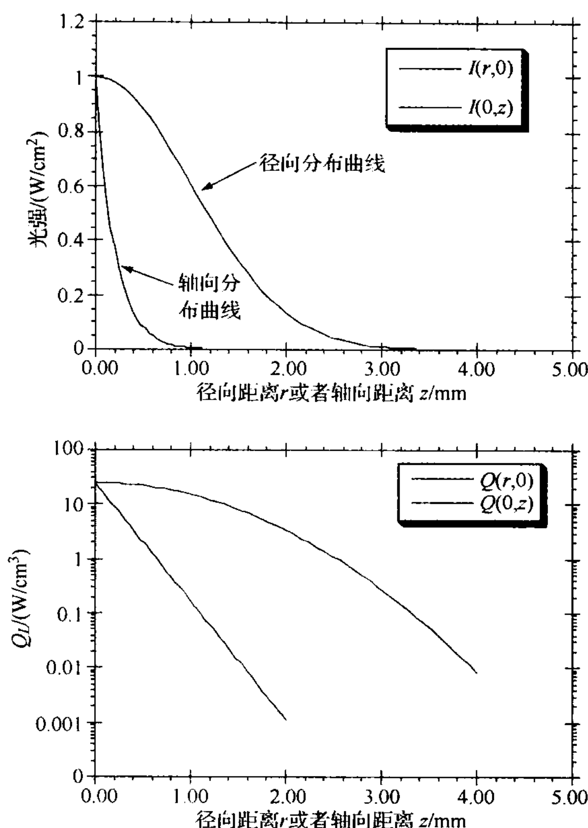
其中 $f(r)$ 是轴对称光束的径向辐射分布。高斯分布是常用的激光照射模型，其分布方程为

$$f(r)=\exp\left(-\frac{2r^2}{\omega_0^2}\right) \tag{17.33}$$

式中 ω_0——光束的“$1/e^2$ 半径”，因为在 $r=\omega_0$ 处，有 $f(r)=1/e^2$。

例 17.4 对于波长为 2.1μm 的高斯分布激光束，假设吸收系数为 $25\mathrm{cm}^{-1}$，散射忽略不计，那么，利用上述方程可以计算出光强的径向分布、生物组织表面 $z=0$ 处的吸收率（即产热率）以及产热率沿光束中心轴 $r=0$ 的轴向分布。请作图显示 r 从 $-2\omega_0$ 到 $2\omega_0$ 范围内的 $I(r,\ 0)$ 曲线，$z=0$ 到 $z=5/\mu_a$ 范围内的光强曲线 $I(0,\ z)$，以及相应的吸收率曲线 $Q_L(r,\ 0)$ 和 $Q_L(0,\ z)$。为了简化，假设 $I_0=1\mathrm{W/cm}^2$。

解： 下图所示是纯吸收生物组织中光强 I 和容积吸收率 Q_L 的径向和轴向分布曲线。



2. 既有散射又有吸收的介质

传输理论的基本参量就是总的光辐射强度 L。对于沿着单位矢量 $\hat{s}$ 所表示的方向传播的光，L 是位置 $\vec{r}$ 的函数，其单位是 $\mathrm{W \cdot cm^{-2} \cdot s^{-1}}$。$L$ 的传输方程可以写为

$$\frac{\mathrm{d}L(\vec{r},\ \hat{s})}{\mathrm{d}s}=-(\mu_a+\mu_s)L(\vec{r},\ \hat{s})+\mu_s\int_{4\pi}p(\hat{s},\hat{s}')L(\vec{r},\hat{s}')\mathrm{d}\omega'+S(\vec{r},\hat{s}) \quad (17.34)$$

此方程表示了光的散射 μ_s 和光的吸收 μ_a 引起的 L 的减小，以及来自其他方向 $\hat{s}'$ 的散射所引起的 L 的增加（见图 17.7）。光的散射系数与吸收系数之和，即 $\mu_t=\mu_s+\mu_a$，被定义为衰减系数。函数 $p(\hat{s},\hat{s}')$ 称为“相函数”，它与光粒子的散射幅度有关，散射幅度用散射角概率分布的标量来表征。注意，如果不考虑散射，并假设光源是准直平行光，那么，式（17.34）就可以简化为上节推导的式（17.29）。另外，式（17.34）中的 $S(\vec{r},\ \hat{s})$ 表示光源函数，可以是物体表面受到的光照，也可以是物体内部产生的荧光，

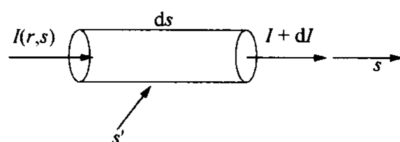


图 17.7　传输理论所描述的光强变化量（其中包括了来自其他方向的散射光 s'）

即内部容积光源。

传输方程是一个含积分的微分方程，一般没有解析解。不过，有几种近似求解方法，如：二流模型（two-flux model）、N流模型（multiflux model）、离散坐标有限元法（the discrete ordinate finite element method）、球谐函数法（the spherical harmonic method）、漫射近似法（the diffusion approximation）以及Monte Carlo模型等。这些方法都有一定的限制条件和假设。本章下面重点讲述比较常用的漫射近似法。

3. 漫射近似法

漫射近似方程是一个二阶微分方程。在散射“大于”吸收的假设条件下，由辐射传输方程（17.34）可以推导出以下漫射近似的微分方程：

$$\nabla^2\phi_d - 3\mu_a\mu_{t'}\phi_d = -\frac{\mu_{s'}}{D}\phi_c \tag{17.35}$$

式中 ϕ_d——漫射光能流率。

方程的参数分别是：$\mu_{s'}=\mu_s(1-g)$，$\mu_{t'}=\mu_a+\mu_{s'}$，$D=1/3\mu_{t'}$，其中 g 是介质的散射各向异性因子，$\mu_{s'}$ 为约化散射系数。此漫射近似方程的解是研究混浊介质中光的分布的有用工具。

物体上某个点的总光能流率 ϕ（单位是 W/cm^2）等于准直光的光能流率 ϕ_c 与漫散射光的光能流率 ϕ_d 之和，即

$$\phi(r,\ z)=\phi_c(r,\ z)+\phi_d(r,\ z) \tag{17.36}$$

ϕ 的大小等于介质空间中微小球体所接受的总的光能除以该球体的表面积。总光能流率是激光与生物组织之间相互作用的关键因素。

上式中准直光的光能流率为

$$\phi_c=I_0(r)(1-r_{sp})\exp(-\mu_{t'}z) \tag{17.37}$$

式中 I_0——入射光的表面光通量密度，单位为 W/cm^2；

r_{sp}——反射系数。

光轴 $r=0$ 处的光学边界条件是

$$\left.\frac{\partial\phi_d}{\partial r}\right|_{r=0}=0 \tag{17.38}$$

其他地方的边界条件是

$$\phi_d-2AD\nabla\phi_d\cdot\tilde{n}=0 \tag{17.39}$$

式中 A——内反射系数；

$\tilde{n}$——指向介质内部的单位法向矢量。

内反射系数 A 表示边界与周围介质之间的折射率不匹配效应，其计算公式为

$$A=\frac{1+r_i}{1-r_i} \tag{17.40}$$

其中 r_i 由如下经验公式估计

$$r_i=-1.440n_{rel}^{-2}+0.710n_{rel}^{-1}+0.688+0.0636n_{rel} \tag{17.41}$$

式中 n_{rel}——生物组织折射率与周围介质折射率之比。

当边界匹配时，$n_{rel}=1$，内反射系数 A 降为 1。

式（17.35）～式（17.41）就是漫射近似的控制微分方程组及其边界条件。对于生物组织等同时具有散射和吸收作用的介质，其光能流率就可以用这组方程的解析解或者数值解来计算和分析。

例如，假设无限大介质中有一个各向同性的点光源，那么，在介质中，与光源相距很"远"的距离 r 处，检测器所测得的光能流率，就可以用上述方程组的 Green' s 解函数来计算，即

$$\phi(r)=\frac{\phi_o}{4\pi D}\frac{e^{-r/\delta}}{r} \tag{17.42}$$

式中　δ——穿透深度，$\delta=\sqrt{D/\mu_a}$。

例 17.5　假设需要测量某散射系数已知的介质中某种吸收物的浓度。如果约化散射系数 $\mu_{s'}$ 已知，并且与各向同性光源相距 r_o 处的相对光强 $\phi(r_o)/\phi_o$ 可以测得，那么，只要 r_o 足够大，能够使漫射近似方程成立，就可以用漫射近似法求得描述介质吸收系数的代数方程。请推导此方程。

解：给定约化散射系数 $\mu_{s'}$，以及 r_o 和 $\phi(r_o)/\phi_o$，应用漫射近似法推导吸收系数 μ_a 的过程如下：

$$\phi(r)=\phi_o\exp(-r/\delta)/(4\pi Dr)$$

$$\log[\phi(r)/\phi_o]=-r/\delta-[\log4+\log\pi+\log D+\log r] \tag{ex. 1}$$

$$D=1/3\mu_{t'}$$

其中 $\delta=\sqrt{D/\mu_a}$。式（ex. 1）左边的值已知，设为 k_1，则

$$k_1=-r/\delta-[\log4+\log\pi+\log D+\log r]$$

$$k_1+\log4+\log\pi+\log r=-r/\delta-\log D \tag{ex. 2}$$

式（ex. 2）左边的值也是一个已知数，设为 k_2，则

$$k_2=-r/\sqrt{D/\mu_a}-\log D$$

$$k_2=-r\sqrt{\mu_a/D}-\log D$$

$$k_2=-r(\sqrt{\mu_a/(1/3\mu_{t'})})-\log(1/3\mu_{t'})$$

$$k_2=-r(\sqrt{3\mu_a\mu_{t'}})-\log(1/3\mu_{t'})$$

$$k_2=-r(\sqrt{3\mu_a(\mu_a+\mu_{s'})})-\log(1/3(\mu_a+\mu_{s'})) \tag{ex. 3}$$

求解上式（ex. 3）就可以得到吸收系数 μ_a。■

17.2.3　光学特性的测量

生物组织的吸收系数 μ_a、散射系数 μ_s 和散射各向异性因子 g 等光学参数的检测一直是生物光学的核心问题。这些数据的测量对于光学在医学诊断和治疗中的应用非常重要。例如，为了准确估算光动力学疗法、光凝法和组织消融等光照治疗过程中组织所接受的局部光通量，就必须知道这些光学参数。并且，生物组织的光学参数不仅对于光学成像和荧光光谱等在体诊断技术具有重要的意义，而且它们本身还

可以提供人体的新陈代谢信息，用于疾病的诊断。

目前，测量生物组织光学特性参数的方法已有很多种。例如，应用准直光透射技术可以测量光的吸收和散射作用之和，即总作用系数 $\mu_a+\mu_s$。测量时，用准直光束照射组织切片，然后通过小孔检测非散射的透射光。小孔可以遮挡并去除散射光。根据比尔-朗伯定律，也就是式（17.30）的扩展形式，就可以计算出被测物的特性参数。适用于散射介质的比尔-朗伯定律是 $I(Z)=I_0\exp[-(\mu_a+\mu_s)z]$，其中 $I(z)$ 是穿过深度 z 之后的未散射的透射光强度。在准直光透射测量中，I_0、$I(z)$ 和 z 均可以测得，因此，就可以计算出 $\mu_a+\mu_s$ 的值。

例 17.6 一束 5mW 的准直激光穿过 4cm 厚的无吸收的散射介质，通过小孔测得的准直透射光为 0.035mW。请计算介质的散射系数。

解：根据比尔-朗伯定律，有

$$\mu_s=\ln(5/0.035)/4=1.2\text{cm}^{-1}$$ ■

另一种生物组织光学特性测量方法是积分球测量法，它可能是最常用的方法。积分球各有一个入口和一个出口，其内壁涂有一层漫反射材料。测量时，将组织切片夹在两个积分球之间，用准直光照射组织样本，通过漫反射光和透射光的积分，就可以同时分别测得组织的反射系数和透射系数。然后，再根据有关模型，用这两个测得的数据计算出吸收系数 μ_a 和约化散射系数 μ_s'。所用的模型可以根据倍增法、δ-Eddington 近似法和 Monte Carlo 法等光子传输理论来建立。

还有一种测量方法是垂直入射反射摄像仪（normal incidence video reflectrometry）。测量时，准直光垂直入射到生物组织上，用 CCD 摄像机或者光纤束采集漫反射光强的空间分布。然后，再根据漫射理论拟合所测得的漫反射空间分布，计算出各个光学特性参数。除非能够通过漫反射分布数据测得总的漫反射光强，否则测得的漫反射空间分布数据只能是绝对值。而绝对值的校准比较困难，因此，这种方法并不适用于临床。

另外，用时间分辨光谱技术和频域技术也可以测量介质的光学特性参数。不过，这些方法所需的仪器设备比较昂贵，对于非科研应用，可能费用太高。

最近出现的基于光纤的斜入射反射仪是一种应用前景较好的在体光学特性检测方法。只要被测样本可以看做半无限大混浊的散射介质（这对于一般在体生物组织都成立），使用这种方法就能够简单准确地测得组织的吸收系数 μ_a 和约化散射系数 μ_s'。本节下面就介绍这种方法。

斜入射光所产生的漫反射空间分布，其中心点与光的入射点之间存在一个位移，这个位移直接与介质的漫射系数 D 相关。通过光纤探头可以将斜入射光传送给被测样品，同时采集样品漫反射空间分布的相对值，这种检测方法不必测量绝对值。然后，根据漫反射分布数据，就可以测得漫射系数 D，并且通过曲线拟合计算有效衰减系数 μ_{eff}，再计算 μ_a 和 μ_s'。这里，有效衰减系数 μ_{eff} 的定义是

$$\mu_{eff}=\sqrt{\mu_a/D} \tag{17.43}$$

以前已经有人用两个各向同性的点源，建立了垂直入射光漫反射的空间分布模

型。一个是正点源，位于生物组织表面下方 $1mfp'$ 处；另一个是负像源，位于组织表面上方。正点源表示组织内部的单个散射源，其像源在 z 方向上的高度与边界条件相关。光子的输运平均自由程（Mean Free Path，MFP）定义为

$$mfp' = 1/(\mu_a + \mu_{s'}) \tag{17.44}$$

对于斜入射光，隐藏的点源应该位于组织内相同光程处，不过此时是在根据 Snell 折射定理确定的新光路上。假设光的入射角已知；生物组织和传递光的某种介质的折射率都已知，那么，就可以得到两个点源的位置变化数据，特别是 x 方向的位移。图 17.8a 和图 17.8b 分别显示了垂直入射光和斜入射光的模型。

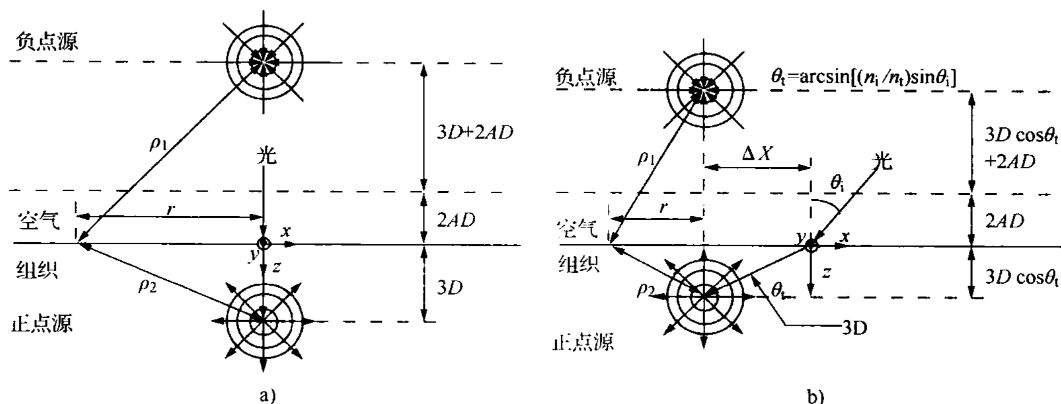


图 17.8　垂直入射光漫射理论模型两个点源的位置（图 a）和斜入射光漫射理论模型两个点源的位置（图 b）

注：图中 y 轴指向纸外，ρ_1 和 ρ_2 分别是正点源和负点源与组织表面检测点之间的距离，检测点与两个点源连轴之间的距离为 r。θ_t 由 Shell 折射定理计算得到。

斜入射光的漫反射分布以两个点源的位置为中心，因此，位移 Δx 就是漫反射分布中心相对于光入射点的距离。

双点源模型的表达式如下：

$$R(x) = \frac{1}{4\pi}\left[\frac{\Delta z(1+\mu_{\text{eff}}\rho_1)\exp(-\mu_{\text{eff}}\rho_1)}{\rho_1^3} + \frac{(\Delta z+2z_b)(1+\mu_{\text{eff}}\rho_2)\exp(-\mu_{\text{eff}}\rho_2)}{\rho_2^3}\right] \tag{17.45}$$

其中 $\Delta z = 3D\cos\theta_{\text{组织}}$。此式可以随意按比例缩放，用于拟合非绝对值的相对漫反射分布曲线。有效衰减系数 μ_{eff} 前面已有定义，ρ_1 和 ρ_2 分别为两个点源与光线采集点（见图 17.8）之间的距离。如图 17.8b 所示，漫射系数 D 可以由 Δx 计算得到：

$$D = \Delta x/(3\sin\theta_{\text{组织}}) \tag{17.46}$$

生物组织的光学特性参数与组织的种类和光的波长有关。例如，对于氩激光等光源发出的绿色光，红色的肝的吸收系数要比黄褐色的鸡胸等组织的吸收系数大得多。可见光和近红外光的一组典型光学特性参数是：吸收系数为 0.1cm^{-1}，散射系数为 100cm^{-1}，各向异性因子为 0.9。在紫外光区域，生物组织中的蛋白质的光吸收为主；在红外光区域，则是水的光吸收为主。近红外（～800nm）区域由于吸收最小且

散射也相当弱，因此被作为诊断检测区域。

光学特性参数检测的一项重要应用是测定血红蛋白氧饱和度，它是一个重要的生理参数。由于氧合血红蛋白和还原血红蛋白具有不同的吸收光谱，因此，通过检测组织的光学特性参数，就可以计算出两种血红蛋白的相对浓度之比。

17.3 光的物理作用及其检测

由以上 17.2 节的内容可知，人体多数组织对光是不透明的，也就是组织对光具有吸收和散射作用。本节讲述光与组织之间的物理作用及其在医学检测和治疗中的应用。所述内容将围绕 4 种基本物理量的变化，即热变化、压力变化、速度变化以及样本的光程变化。其中，利用光既可以诱发也可以测量组织样本的热变化和压力变化，通过光的多普勒频移现象可以检测速度变化，通过两束以上光束的干涉图案则可以测量样本光程长度的变化。

17.3.1 热作用和温度监测

人体所有组织对于不同波长的光都有吸收作用，光被吸收之后，光能就转化为热能。同时，与任何热力学温度在 0 度以上的物体一样，人体组织还会产生光辐射，称为黑体辐射。因此，光可以加热人体组织，用于治疗；反过来，通过检测组织的光辐射，还可以测量组织的温度。

1. 温度监测

一般，使用既廉价又可靠的热敏电阻或者热电偶等电子传感器，就可以检测温度。不过，如果存在较强的电磁场干扰，这些传感器就不能用，这种情况在医院里经常会遇到。而且，这些传感器也不能用于监测激光照射所引起的温度上升，因为它们本身会吸收激光辐射。因此，人们开发了几种间接测量人体温度的光学方法。例如，液晶光极（liquid crystal optrode）会随着温度的变化产生明显的颜色变化；干涉仪传感器则可以通过其相位和干涉条纹的变化，反映出温度变化引起的光程长度变化；另外，很多物体的发光与温度之间存在很强的相关性，也可以用于温度测量。

下面先介绍一种直接测量温度的传感器，就是所谓的热像仪（或称辐射计）。这种系统最初由军队研制并使用，可以在黑暗环境中探测交通工具、人体、船只等。温度为 T 的黑体的总辐射量计算公式是

$$I=\varepsilon\sigma T^4 \tag{17.47}$$

式中 ε——辐射系数；

σ——Stefan-Boltzmann 常数。

室温下，物体的辐射主要属于光谱上的远红外区域。随着温度的上升，辐射会逐渐出现在近红外区域，最后到达可见光区域。生物组织的辐射系数为 1，即 $\varepsilon=1$，因此，其总辐射只与温度 T 相关。在军事上，用红外辐射产生的热像图可以显示出物体的温度分布，同理，这种技术可以在医学上用于监测人体的体表温度。例如，

配有视频显示器的热像仪可用于显示人体胸部的温度分布。由于乳腺癌组织的温度比健康组织的温度高，因此，从这种图像上就可以观察乳腺癌的病情。注意，在乳腺癌筛查中，虽然红外成像不如乳房 X 射线透射成像可靠，但是，这种方法对人体毫无损害，病人不会受到辐射照射。不过，用红外热像仪测量时，被测物体表面到探测器之间必须存在直接的视线通路，不能有遮挡。如果有其他物体挡住了视线，那么，有时可以在探测器上连接红外线光纤，通过光纤传导红外辐射。这种光纤辐射计已经被提议用于生物组织的微波热疗和激光热疗。

2. 光致发热

如前所述，除了监测人体的光辐射之外，光本身还可用于加热具有光吸收性的生物组织。生物组织的光致发热有很多应用，包括生物刺激、封口和血管焊接、组织损毁和组织汽化等。生物刺激是指光诱导的生物组织短时间加热，这种方法可以刺激神经，加速创口的愈合。另外，具有高吸收特性的激光可以促进组织之间的融合，特别是血管的连接；也可以促使血液凝固，用于手术过程中的止血。如果光致发热使生物组织的温度上升到 45℃以上，那么，就会发生组织坏死，这可用于癌症和前列腺组织增生的治疗。如果照射光的功率密度再增加，那么，生物组织就会发生消融和汽化，所谓的矫正视力的激光角膜切除术就是这种情况。

光致发热一般都使用激光，可以选择紫外到红外区域的各种不同波长的激光，激光的功率密度（即光束功率与照射面积之比）以及持续时间都可以调节。通过改变功率密度或者激光脉冲的持续时间，可以改变传递给生物组织的能量，从而控制温度的上升。激光的功率密度较高时，组织会发生凝固、坏死和汽化；而功率密度较低，只产生较小的加热效应。组织吸收的能量是照射光波长的函数。对于某些波长，如那些水的强吸收波段，由于生物组织的主要成分是水，因此，组织对于这些波段的激光的吸收很强。如果用这些波长的光进行照射，光的能量会立即被靠近表面的组织薄层大量吸收，使该组织层迅速升温。激光角膜切除手术就是使用这种机理。如果用其他波长的光进行照射，生物组织的光吸收较小，加热的速度较慢，加热的深度就会增加，可以加热较大体积的生物组织，前列腺手术就是这种情况。

3. 温度上升与光子吸收率之间的关系

光与物体相互作用的所产生的热力学不可逆反应就是光子能量被物质相吸收的过程。在不存在热传导的情况下，生物组织温度的上升由热力学状态方程决定。状态方程表明物质内部能量的变化与温度的上升成正比。在无热传导和其他热传输作用的前提下，物质内部能量的变化就等于物质的光子吸收率，它用时间微商表示，即

$$\Delta U/\Delta t=\rho C\Delta T/\Delta t\approx Q_L \tag{17.48}$$

式中　ΔT——温度的上升量，单位为 K；

Q_L——组织的单位体积光子吸收率，单位为 W/m^3；

ρ——物体的质量密度，单位为 kg/m^3；

C——比热容，单位为 $J/(kg\cdot K)$；

ρC——单位体积的热容量，单位为 $J/(m^3 \cdot K)$；

Δt——光照持续时间，即曝光时间。

光子吸收率 Q_L 是物质利用光子升温的重要因素，它就是光源函数项，也称为能量沉积率。在一般光作用过程中，物体的光子吸收率与辐照度成正比，比例常数就是吸收系数，即

$$Q_L=\mu_a\phi \tag{17.49}$$

辐照度 ϕ 与组织中光子的空间分布有关⊖。例如，对于纯吸收物质，应用比尔-朗伯定律，并假设光束具有高斯分布，则

$$Q_L=\mu_a\times\phi_0\exp(-\mu_a z)\exp(-2r^2/\omega^2) \tag{17.50}$$

式中 ϕ_0——入射光强度；

z——深度；

r——径向距离。

如果存在光的散射作用，那么，可以用前面所述的各种散射模型来表示 ϕ_0。

式（17.48）成立的假设条件是不存在传导、对流等其他热作用过程。如果这些过程可以忽略不计，例如，持续时间非常短的激光脉冲就是如此，那么，由式（17.48）可知，物体温度的上升可以计算如下

$$\Delta T\approx Q_L\Delta t/\rho C \tag{17.51}$$

式中 Δt——光照的持续时间。

17.5 节将进一步介绍激光与组织之间相互作用所产生的其他热效应。为了使照射过程中的热扩散可以忽略不计，式（17.51）只有在非常短的脉冲照射下才能成立，照射脉冲时间宽度的上限临界值可用下式计算：

$$\Delta t_{\max}=\frac{1}{\sqrt{4\mu_a^2\alpha}} \tag{17.52}$$

式中 α——物体的热扩散系数，水的 α 值约为 $1400cm^2/s$。

17.3.2 激光多普勒测速仪

除了使物体升温以外，光与运动物体之间的物理作用还可以产生多普勒现象，也就是运动物体的速度所产生的频移。例如，人体血管中流动的直径约为 7 μm 的血红细胞就是一种运动物体，利用多普勒技术测量血流速度的方法已经形成多种临床应用，包括心脏监护、冠状动脉血管成形术、冠状动脉狭窄检测、体表组织血流测量以及分娩过程中胎儿头皮的血流监测等。

当频率为 f、速度为 c 的入射光照射在静止物体上时，其反射光的频率保持不变。但是，如果被照射物体以速度 v 运动时，那么，反射光的频率 f' 就与原入射光频率 f 不同，两个频率之差（即频移）δf 称为多普勒频移，表示为

$$\delta f=f-f' \tag{17.53}$$

⊖ 对于多光子的光与组织作用过程，ϕ 的指数成倍增加，如双光子过程的光子吸收率为 $Q_L=\mu_a*\phi^2$。——原文注

光束照射在流动的血红细胞上所产生的多普勒频移可以用式（17.54）计算：

$$\delta f/f=2vn\cos\theta/c_0 \tag{17.54}$$

式中　v——血红细胞流动速度；

n——介质的折射系数；

c_0——生物组织中的光速；

f——入射光频率；

θ——入射光与血管之间形成的夹角。

由于波长 λ 等于光速除以频率，因此，上式可以变为

$$\delta f=2vn\cos\theta/\lambda_o \tag{17.55}$$

在声波中，多普勒效应可以用驶向观察者的火车来说明，当火车越来越接近观察者时，火车的鸣笛声听起来音调会增高。

例 17.7　用激光多普勒系统测量血流速度。假设使用波长为 514nm 的氩激光，探头与血管成 30°角度，血液的折射系数为 1.33，测得的频移为 60kHz。请计算血流速度，并说明该血流速度是否符合实际生理情况。如果探头与被测血管之间的角度有 10°误差，那么，对计算结果会有什么影响？

解：利用式（17.55），计算速度如下

$$v=(\delta f\cdot\lambda)/(2n\cos(\theta))=[(60\times10^3)(514\times10^{-9})]/[2(\cos(30))(1.33)]=1.34\text{cm/s}$$

对于人体内中等大小直径的血管而言，这个速度是合理的。如果探头角度有 10°误差，比如将 30°误作为 40°来计算，那么得到的速度为 1.51cm/s，误差很大，达 13%。■

虽然激光多普勒是一种很好的监测速度的方法，但是，它在临床中的应用并没有得到推广，原因之一是因为临床上对于诊断有价值的物理量是血流的流量（单位为 cm^3/s），相当于血液中各种颗粒的平均流速，但这种流量并不能用多普勒方法直接测得，而是要通过计算才能得到。如果刚性管道中的水流速度已知，那么，计算其流量就很简单。不过，仅仅根据多普勒测得的血流速度来计算血液的流量，却是一个很困难的问题。特别是对于狭小的血管，由于血液比水粘稠得多，血流的特性很复杂。而且，血管并不是刚性直管，其中的血流是脉动式的。最后，测量血流时，为了获得一系列速度数据，用于流量计算，经常需要将光纤探头插入血管中进行测量。在这种情况下，探头本身也会对血流的流量产生影响。

总之，激光多普勒测速是一种简单的方法，通过设计，可以用于测量物质流量的相对变化，但是，要用这种方法测量绝对流量却很难。因此，临床上广泛使用的血流量标准测量法是热稀释法。该方法的原理是：在血液循环系统的某个部位注入一定量的低温溶液，引入一个预定的血液热容量变化，然后在血流下游某处血液温度已经混合均匀的地方，检测血流的温度变化，从而计算出血流量。参见本书 9.3.3 节。

17.3.3　干涉仪

干涉现象也是光的一种物理作用，通常由来自同一光源的两束或多束光波叠加

产生。光波由电场和磁场的振荡构成，而电场和磁场都是矢量，它们的叠加是矢量相加。因此，如果来自同一光源的两束或多束光波分别通过不同的传播路径，然后，再重新组合在一起，那么，这些光波之间就会发生增强或者抵消的干涉效应。空间交替出现的增强和抵消干涉效应表现为一系列干涉条纹。如果其中某条光束传播路径受到温度、压力或者折射率等的微小变动的影响，发生了变化，那么，从这条路径过来的光束与无变化的参考光束重新相遇时，所形成的干涉条纹就会产生移位。这种移位可以通过光电技术测出，精度可达条纹间距的 10^{-4}。利用这种干涉方法，可以非常精确地测量与目标参量变化相关的有用信息。光传播路径变化的测量精度可以达到 1%波长的数量级，对于可见光，相当于 5×10^{-9}m。

干涉仪有很多种，最早的一种是瑞利（Rayleigh）折射计，后来演变为 Mach-Zehnder 干涉仪。为了简明扼要，本节主要介绍基于 Rayleigh 原理的两种比较精密的干涉仪：双光束 Michelson 干涉仪和多光束 Fabr-Perot 干涉仪。如图 17.9 所示，在 Michelson 干涉仪中，光源发出的光被分光镜或者光纤分为两束，当这两束光分别从两个镀银全反射镜反射回来时，分光镜或光纤又会将它们重新组合起来。干涉仪中有时还配有一块补偿板，可用于调节光路，使测量前或者仪器空载时，测量光路和参考光路的长度相等。被测的变动可以是压力或者温度变化引起的物体应变，从而使样品光路上的路径长度发生变化；也可以是光路上加入生物组织样本所引起的变化；或者是用组织样本代替反射镜之后产生的变化。例如，用 Michelson 干涉仪可以测量生物组织的厚度，特别是在矫正视力的激光角膜切除手术过程中，可用于监测角膜的厚度。配上低相干光源之后，这种干涉仪还可用于各种体表组织的光学相干断层成像（optical coherence tomographic imaging）。当两束干涉光的幅值相等时，以光轴为中心的干涉条纹圈的控制方程是

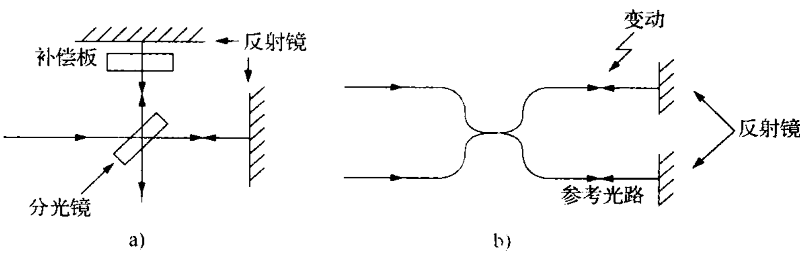


图 17.9 透镜（图 a）和光纤（图 b）结构的 Michelson 干涉仪

$$I=4I_0\cos^2(\delta/2) \tag{17.56}$$

其中 I_0 为入射光强度，相位差 δ 定义为

$$\delta=2\pi\Delta/\lambda \tag{17.57}$$

其中 λ 为波长，两条光路光程的净差 Δ 定义为

$$\Delta=2d\cos(\theta)+\lambda/2=(m+0.5)\lambda \tag{17.58}$$

显然，由此式可以推导出以下方程

$$2d\cos(\theta)=m\lambda \tag{17.59}$$

式中 $2d$——从分光镜开始的两条光路光程之差；

m——干涉条纹数目；

θ——入射角，$\theta=0$ 时是垂直入射。

注意，光程与实际物理距离是不同的。如果将折射率为常数的一块薄板、一小瓶气体或者一片吸收和散射极小的组织薄片，插入其中一条光路中，那么其光程 d 为

$$d=(n_s-n_{空气})L$$

式中 L——被测物的实际厚度；

n_s——被测物的折射率；

$n_{空气}$——空气的折射率。

例 17.8 将一片 $n\approx1.33$ 的清澈组织薄片，比如一小块眼睛角膜，垂直插入 Michelson 干涉仪的一路光线中，结果使干涉条纹移动了 50 个条纹的距离。假设光的波长为 589nm，请计算这块组织的厚度。

解： 由式（17.59）的 $2d\cos(\theta)=m\lambda$ 可得 $d=(50\times0.589\ \mu m)/2=14.72\ \mu m$。此计算值是光程长度，生物组织的实际物理厚度还必须利用样本和空气的折射率来计算，其值为 $L=d/(n_s-n_{空气})=14.72\mu m/(1.33-1.0)=44.6\mu m$。 ■

Michelson 干涉仪和 Mach-Zenhnder 干涉仪都属于双光束干涉仪，这类干涉仪的一个缺点是，其干涉条纹亮度呈现式（17.59）所示的正弦变化，最大值或者最小值的精确位点难以分辨，因此，其测量精度受到限制。对于测量仪精度要求非常高的仪器，如精密的光谱仪，这个缺点就会产生较严重的影响。利用多光束干涉法，可以克服这个缺点，大大提高测量精度。图 17.10 所示的 Fabry-Perot 干涉仪使用的就是多光束。这种干涉仪用两块平行板产生多束透射光形成干涉条纹。两块平行板一般是封装在一起的厚玻璃或者石英，中间夹着空气层。平行板内表面的平整度和反射率很重要，一般抛光精度要高于 $\lambda/50$，并且需要镀上高反射银膜或铝膜。对于 400nm以上波长的可见光，镀银比较好；而对于 400nm 以下波长的光则铝的反射率比银高。镀层必须很薄，才能使其具有较好的透射性，例如，银质镀层的厚度一般约为 50nm。在很多情况下，平行板的外表面并不是与内表面完全平行，而是特意做成与内表面之间形成几分小角度，以便消除玻璃板本身引起的干涉条纹伪迹，因为两面平行的玻璃板自身就是一个平行板干涉仪，会产生干涉条纹伪迹。如果两块平行板之间的间距 L 固定，那么，干涉仪称为 Fabry-Perot 标准具。仪器中，各个相邻平行光束的光程差决定了某个空间位置 P 处光束叠加的结果，光程差的计算公式是

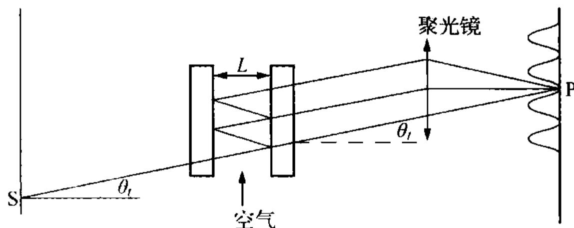


图 17.10 Fabry-Perot 多光束干涉仪

$$\Delta=2n_f L\cos\theta_t=m\lambda_{\min}=2L\cos\theta_t \tag{17.60}$$

对于空气，$n_f=1$。

进入平行板的来自光源不同点的其他光束，如果与光轴形成相同的倾斜角 θ_t，光程差相等，这些光束也会聚在 P 点。如果 L 固定不变，是常数，那么，以上公式就表示角度为 θ_t 的入射光束之间的光程差 Δ，如图 17.11a 和图 17.11b 所示，这些等倾光束所形成的聚焦干涉条纹是同心圆环。如果间距 L 是随时间变化的变量，那么，检测器 D 就会记录到如图 17.11c 所示的随时间变化的干涉曲线。

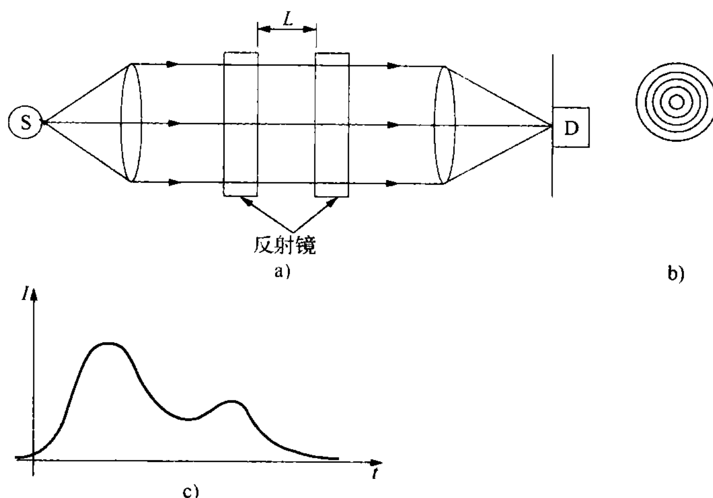


图 17.11　Fabry-Perot 干涉仪（图 a）及其随空间（图 b）和随时间（图 c）变化的干涉条纹

Fabry-Perot 干涉仪干涉条纹的亮度变化，也就是所谓的干涉条纹分布，是光程差（即相位）的函数。条纹的锐度是决定仪器分辨能力的重要因素。利用三角恒等式 $\cos\delta=1-2\sin^2(\delta/2)$，可得光束的透射率计算公式为

$$T=I_T/I_i=1/\{(1+[4r^2/(1-r^2)^2]\sin^2(\delta/2)\} \tag{17.61}$$

式中　r——反射系数；

　　δ——相位差。

分母方括号内的项被称为精细度 F；于是

$$T=1/[1+F\sin^2(\delta/2)] \tag{17.62}$$

注意，由于 $0<r<1$，因此 $0<F<\infty$。精细度 F 也是一种干涉条纹对比度的衡量指标，即

$$F=[(I_T)_{\max}-(I_T)_{\min}]/I_T\min=(T_{\max}-T_{\min})/T_{\min} \tag{17.63}$$

注意，当 $\sin(\delta/2)=0$ 时，有 $T_{\max}=1$；当 $\sin(\delta/2)=\pm1$ 时，有 $T_{\min}=1/(1+F)$。不同 r 值下干涉条纹的亮度分布曲线如图 17.12 所示。

由图中可见，当 $\delta=m2\pi$ 时，$T=T_{\max}=1$；当 $\delta=(m+1/2)2\pi$ 时；$T=T_{\min}=1/(1+F)$。并且，$T_{\max}=1$ 与 r 无关，而 $T_{\min}$ 不会等于零，只是当 r 趋向于 1 时 $T_{\min}$ 趋

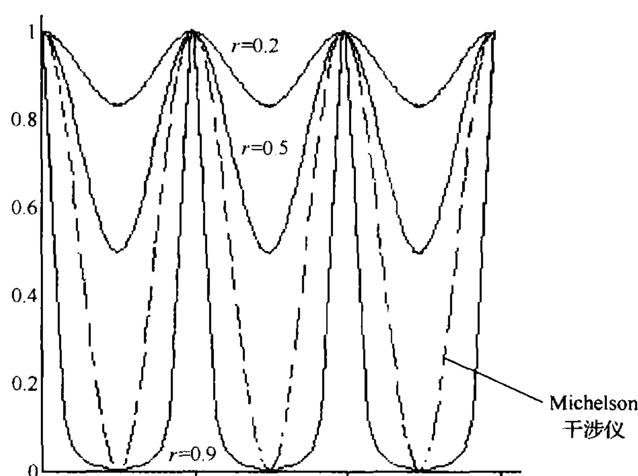


图 17.12　具有不同反射率的多光束 Fabry-Ferot 干涉仪与双光束 Michelson 干涉仪之间分辨率的比较

向于 0。另外，r 值越大，干涉条纹的波峰就越尖锐，条纹之间的大部分区域干涉强度都接近于零。注意，图中虚线所示的是 Michelson 干涉仪的分布曲线，为了便于比较，图中显示的都是按照最大值归一化之后的数据。可见，由于 Michelson 干涉条纹的亮度分布直接与 $\cos^2(\delta/2)$ 相关，因此，其条纹不够尖锐。Fabry-Petot 光纤颅内压监测仪是干涉仪在临床上的一种应用，用于诊断严重的脑损伤或脑积水，脑积水是指脑室和蛛网膜下腔脑脊液的增多。

17.4　生化物质的光学测量技术

近年来，许多医学仪器公司和高校都投入极大的热情，积极开发各种临床光学检测技术，用于癌症的检测，以及血糖、乳酸、白蛋白和胆固醇等血液生化指标的定量监测等。人们最熟悉的临床光学监测技术可能就是血氧传感器，它可以根据氧合血红蛋白和还原血红蛋白的光谱吸收峰，定量检测人体中这两种物质的变化量。医学检测和诊断中常用的光学原理包括光的吸收、散射、荧光发射和光的偏振等，检测的主要参量是光穿过介质之后的光强变化，这种光强变化与入射光波长以及偏振性有关。当被测物浓度适中时，光的吸收或者荧光发射所引起的光强变化与物质浓度变化之间接近线性关系；但是，当被测物浓度很高时，两者就呈现非线性关系。

17.4.1　基于光吸收原理的分光光度计

很多研究人员曾经试图利用红外线吸收进行血液生化物质等临床检测，尤其是血糖的监测和癌症诊断。本章前面已经介绍了适用于纯吸收介质和荧光发射的比尔-朗伯定律，也就是式（17.17），将该式表示为对数形式，可以写为

$$A=\ln(T)=\ln(I_0/I)=\mu z=z\Sigma\varepsilon_i C_i \quad (17.64)$$

式中 A——吸光度；

T——透射率；

I_0——入射光强度；

I——透射光强度；

z——光程长度；

μ——吸收系数，也就是被测介质中各个不同组分摩尔吸收系数 ε_i 与浓度 C_i 乘积之和。

对于生物组织和血液，在波长为 2.5～12μm 的中红外（mid-InfraRed，mid-IR）区域，式（17.64）才成立，因为，在这个波长范围内，生物体化学物质的吸收峰比较尖锐明显，并且光的散射很弱。不过，此波长范围内生物组织中水的光吸收比任何血液生化成分的光吸收都要大好几个数量级，因此，如果用红外线测量，样本光程的长度只能很小，仅为微米数量级。这样，就形成了表面测量，也就是，在人体皮肤表面进行测量，或者将光纤插入体腔内，进行腔内表面测量。但是，现有的可用于这个波长范围的光纤都有毒，且吸湿性强，很坚硬。因此，由于存在水的光吸收问题，再加上没有合适的光纤，这种利用中红外光进行在体检测和诊断的方法目前还很难实现。中红外检测系统中，可用的光源有：宽光谱的钨丝、能斯特灯、镍铬合金丝、硅碳棒以及谱带较窄且波长可调的液氮冷却式激光二极管等。光强检测器有：冷却式碲镉汞（Mercury Cadmium Telluride，MCT），热电堆和热敏电阻等。光学器件则有：氯化钠或溴化钾单晶制作的样品池，镀金的反射镜和光栅等。

在近红外（Near-InfraRed，NIR）波长区域，光谱受水的影响不像中红外区域那么大，光程长度可达 1mm～1cm。并且，低 OH 硅质光纤在这个波段的透明度相当好，且无毒、低吸湿。NIR 的波长范围为 700～2500nm，该波段的光吸收主要由 700～1000nm 的低能量电子振动以及 1000～2500nm 的分子化学键振动所产生的谐波组成。含有—CH、—OH 和—NH 等化学键的物质会在这个波段产生谐波，一般，能够检测到的只有分子光谱的一、二和三次谐波，且谱峰较宽。只有高浓度的化学样本的谐波谱峰才能用于定量检测，且谱峰的强度随着谐波次数的增加迅速减小。NIR 吸收波段还会受到温度、压力和氢键的影响，这些影响组合在一起会产生严重的干扰。在 700～1200nm 波段还存在光的散射效应的影响。因此，与 mid-IR 分光光度计不同，NIR 分光光度计主要用于定性观察，并不适用于定量检测。不过，应用多变量统计等方法，也可以用 NIR 光谱进行定量分析。NIR 波段可用的光源有：宽光谱的钨丝、发光二极管 LED、激光汞固体激光器（如波长可调的钛蓝宝石激光器等）以及各种激光二极管。可用的检测器有：适用于 1 μm 波长的硅检测器、适用于 1.7 μm 波长的锗检测器以及适用于 5.5 μm 波长的锑化铟检测器等。光学器件有：低 OH 透镜、石英、玻璃、光栅和反射镜等。

改变光波波长的方法主要有色散和非色散两类。色散型系统使用反射光栅或者透射棱镜来分离不同波长的光。而非色散型系统则用一系列不同波长的滤光器，或

者用红外傅里叶变换器（Fourier Transform InfraRed，FTIR）来分离光谱。在 FTIR 系统中，首先使用类似图 17.9 所示的 Michelson 干涉仪采集完整的光谱数据，然后，再通过傅里叶变换，将以时间为变量的干涉图谱变换为以频率为变量的光谱图。色散和非色散两种方法都可以用于建立覆盖 NIR 和 mid-IR 波段的光谱。色散系统的优点是分辨率高，可以区分波长靠得很近的谱带；而非色散系统由于各种波长的光都通过被测样本之后才被分离，因此，一般具有较高的光通量。

17.4.2 基于散射光的检测技术

光的散射主要有弹性和非弹性两种。弹性散射可以用 Mie 理论来描述。当散射微粒小于波长时也可以用 Rayleigh 散射来描述。Mie 理论表明散射光的强度与溶液浓度、容积大小和散射微粒的形状有关。非弹性散射中散射光子的能量发生变化，可以用 Raman 散射来描述。荧光就是非弹性散射光。

利用散射光进行测量时，通常检测的是反射光的强度。反射光有两种不同的形式：一种是发生在介质表面上的镜面反射。这种反射过程中光波没有进入介质内，反射光并不包含介质材料的任何信息，因此，检测时要想办法尽可能减少或者消除这种镜面反射光。另一种是光波进入介质之后所产生的漫反射。在介质中，部分光波经过多重散射之后，会返回到介质表面。17.2 节所述的光子传输理论可用于描述生物组织中光的漫反射效应，解释漫反射的检测原理。

弹性散射光既可以用于癌症诊断，也可以用于糖尿病患者的非侵入式血糖监测等。在血糖浓度监测应用中，可以使用光强可调的频域 NIR 光谱仪，将物体的约化散射系数与吸收系数分开，以便检测出与人体肌肉中血糖含量相关的约化散射系数。显然，血糖浓度在生理范围内的增加会减少肌肉组织的总散射量，因此，这种方法作为血糖浓度随时间变化的一种相对值测量方法，具有较好的应用前景。但是，弹性散射检测用于物质浓度的监测还是存在难以克服的缺点。其中，特异性是这种方法的最大问题，因为与血糖浓度无关的其他生理效应也会产生类似的约化散射系数的时间变化，并且，与光的吸收不同，光的弹性散射几乎与波长无关。另一个问题是约化散射系数的测量精度不够高，光的散射与吸收之间的分离比较困难。要检测出如此微弱的散射光变化量，同时要避免生物组织（尤其是血红蛋白）较大的光吸收变化的影响，这是很困难的。另外，检测时还必须考虑各种生物组织具有的不同折射率。生物组织的散射由各种物质成分以及质膜、线粒体、细胞核等细胞器产生，它们都具有不同的折射率。血糖浓度及其分布对于散射光的影响是一个非常复杂的问题，只有在细胞水平上搞清楚这个问题，才可能将上述方法用于实际临床检测。由于约化散射系数与细胞密度等因素有关，这类仪器可能需要根据某种标准进行在体校准。最后，还必须考虑可能改变约化散射系数的其他各种因素，如温度、血红细胞浓度、电解质浓度、细胞内外之间水的传输等。作为一种癌症诊断的筛查方法，散射测量在组织切片和离体细胞的检测中具有较好的发展前景。癌症引起的很多生物组织变化是形态上的改变，是细胞和亚细胞结构大小和形状的变化，不是化学物

质的变化。因此，测量时应该使所产生的大部分弹性散射变化与组织形态的变化相关。如果谨慎选择测量所用的入射光波长，将波长选取在水和血红蛋白的主吸收区之外，并且，检测出漫散射光随光线入射角变化的数据，那么，就可能用这种方法辅助癌症的诊断。

近30多年来，使用非弹性Raman散射光谱仪的主要是物理学家和化学家。Raman光谱仪已经成为一种非常重要的研究工具，可用于分析蛋白质、酶、免疫球蛋白、核酸、核蛋白、脂类等各种生物分子，以及生物膜和碳水化合物等。并且，随着高功率激光光源和高灵敏度检测器的发展，Raman光谱仪也正在成为有用的临床检测和诊断工具。当单波长的单色光入射到介质上时，就会产生Raman散射现象。除了透射光之外，一部分光被散射。并且，频率为ω_0的部分入射光会产生$\pm\omega_m$的频移，该频移与物质分子的转动、振动和能级的变化相关。一般，散射光的强度和偏振取决于测量点相对于入射光的位置。多数应用中检测的是频率为$\omega_0-\omega_m$的Stokes散射谱线。此时，Raman光谱仪的检测谱带要移到比激发光波长长的区域。

与红外光谱技术相似，Raman散射光谱也可以用于研究各种分子结构，这两种光谱都可以检测分子极化和偶极矩变化的特征光谱。当两种方法单独使用都不能测出某种分子的所有能态时，可以将Raman散射光谱作为红外吸收光谱的一种补充。不过，某些分子的某些能态，这两种方法可能都测不出。根据偶极子的非简谐振子模型，除了基频振动以外，还存在谐波。Raman光谱的一个优点就是谐波比基波要微弱得多，因此，与吸收型光谱相比，其光谱简单得多。Raman光谱用于生物检测时，还有一个优点是水的Raman散射光谱很弱，对溶质光谱的干扰很小，因此，可用于测量各种水溶液的光谱，这点与红外光谱不同。但是，Raman光谱的信号本身就很弱，因此，直到最近，快速CCD阵列取代了慢速光电倍增管，并且高功率近红外激光二极管产品出现之后，Raman光谱技术才有可能用于鉴别正常和异常生物组织，准实时定量监测血液中的各种生化物质。另外，研究人员还应用偏最小二乘法（Partial Least Squares，PLS）等统计学方法来辅助Raman光谱的生化物质浓度计算。

与弹性散射光谱一样，Raman非弹性散射光谱也已用于临床监测和诊断。诊断方法就是寻找癌变组织等生物组织中存在的，由不同化学物质引起的各种光谱谱峰，以及各个谱峰之间的光强差。定量分析时，只需计算强度差。生物组织测量时，除了水的吸收问题之外，还有一个问题就是高浓度蛋白质等荧光物质的影响。富含血管的生物组织所产生的自发荧光会造成较强的荧光背景信号。由于在波长较长的区域，生物组织的荧光发射较弱，因此，人们开发了使用近红外（NIR）波长激发光的仪器来克服自发荧光的问题。NIR波长激发光还有一个优点，就是在穿过体积较大的组织样本时，与可见光和紫外光等较短波长的光相比，波长较长的NIR光的吸收和散射都比较小。但是，随着波长的增加，在荧光减少的同时，Raman信号的强度也会迅速衰减，它与波长的四次方成反比。因此，既要使荧光最小化，又要保证足够强的Raman信号，只能采取折中的办法。有人提出，在眼睛部位进行Raman光谱检测可以避免自发荧光的影响。不过，为了避免损伤眼睛，这种检测所用的激发光

功率必须很小，这会大大降低信噪比。最后，与红外和近红外吸收型光谱仪一样，进行生物体内低浓度物质的在体定量测定时，还必须解决各种不同化学物质 Raman 光谱的重叠问题。

17.4.3　基于物体发光特性的检测技术

如前所述，发光是物质吸收某个波长的电磁辐射光子，然后在另一个波长上再发射光子的过程。生物组织等介质中的分子吸收光子之后，上升到某个激发态，然后返回到某个较低能态时，分子就会发光。物体的冷发光有荧光和磷光两种。荧光发射过程中的能量转变不涉及电子自旋的变化，光的再发射速度较快。因此，荧光只在激发时才产生，而磷光则在激发停止之后还可以持续一段时间。例如，普通电视机屏幕在开机时会产生荧光，而在关机后的短时间内还会发出磷光。

荧光测量技术已用于临床检测和诊断，它利用生物组织的内源性和外源性荧光来获得诊断信息，特别是用于癌症诊断和动脉血管内壁的沉积物诊断。内源性荧光来自蛋白质、核酸和核苷酸辅酶等物质产生的自发荧光，而外源性荧光则是生物组织摄入荧光染料之后所产生的荧光。外源性荧光已用于监测血糖、细胞内钙离子、蛋白质和核苷酸辅酶等物质。与化学的稀溶液荧光检测不同，生物组织的外源性荧光检测受到内源性自发荧光以及光的散射和吸收等干扰。

荧光传感器检测的是荧光强度，即

$$I_f=\Phi_f(I_0-I) \tag{17.65}$$

式中　I_f——测得的荧光强度；

Φ_f——荧光系数；

I_0——照射在样品上的入射光强度；

I——穿过样品的透射光强度。

荧光系数可以分解成 3 个因子：

1）量子产量（quantum yield），它是激发态分子通过发射光子衰变而不是非辐射性损耗能量的概率。该因子的数值变化范围是 1.00～0.05，并且会在纳秒数量级的时间上发生变化。因此，除了光强测量以外，还可以利用脉冲光源和快速检测器进行时间分辨荧光光谱测量。

2）几何参数，它是与检测器探头结构有关的荧光发射检测角度。

3）检测器本身对于所发射荧光波长的检测效率。

荧光发射是一种吸收和再发射过程，可以用比尔定律描述如下：

$$I_f=\Phi_f I_0[1-\exp(\varepsilon Cl)] \tag{17.66}$$

式中　C——被测物浓度；

l——光程长度；

ε——摩尔吸收系数。

此式可以展开成指数级数，并且，对于弱吸光物质，即 $\varepsilon Cl<0.05$ 时，级数的第一项为主，其他项可以忽略不计。因此，在此条件下，荧光传感器测得的荧光强度

就与被测物的浓度呈线性关系，可以表示为

$$I_f=\Phi_f I_0 \varepsilon C l \tag{17.67}$$

多数荧光传感器检测的是荧光强度，不过，也有一些仪器在时域或频域上检测荧光寿命。为了获得尽可能多的信息，特别是对于科研和教学应用，普通台式荧光计通常要用两个单色器，一般均为光栅式波长分离器；检测器则用光电倍增管(PhotoMultiplier Ttube，PMT）或者CCD阵列；光源是紫外光和可见光为主的宽光谱氙灯。测量时，光源发出的光首先经过第一个单色器，根据用户的设置，该单色器将所需波长的激发光分离出来；然后，激发光穿过被测样品之后，再经过第二个单色器；最后，由光电倍增管扫描采集该单色器光栅反射出来的各个不同波长上的荧光强度，或者，用CCD检测阵列同时采集不同波长的荧光，从而获得荧光强度随时间变化的曲线。CCD阵列的优点是能够实时采集荧光光谱；PMT的优点则是灵敏度较高。很多荧光检测系统中，还将一小部分入射光分离出来，由参考检测器检测，用于校正光源的波动。如果某种特定临床应用（如宫颈癌检测和血糖检测等）的最佳波长设置已经确定，那么，就可以设计一种比较简单稳定的专用荧光检测系统。这种系统不必使用单色器，只要用几个特定波长的滤光片，选取两个或多个离散的工作波长就可以了。另外，还可以用光纤将光传递到远处，或者将远处的光采集回来。由于激发光的波长和发射光的波长不同，用同一条光纤或光纤束就可以同时传出激发光并采回荧光。设计时，最重要的是要考虑光源、所用染料、被测样本和检测器的光谱特征。例如，任何荧光系统的设计都要根据被测生物组织的特性，考虑光的吸收、散射和自发荧光等因素的影响。

17.4.4　基于光偏振特性的测量技术

17.1.2节讲述了有关偏振光的一些基本电磁理论，本节将介绍偏振光的应用。偏振光可用于定量检测血糖等生化物质，或者定性分析生物组织的特性，尤其可用于癌细胞的识别，以及正常眼睛和青光眼的神经纤维层鉴别。

多年来，由光学活性物质引起的线偏振光的旋转已经被用于定量检测溶液的浓度，出现了适用于测定不同光学活性物质的各种偏振计，例如，专用于糖溶液含量检测的旋光糖量计等。这些仪器的检测原理是：当偏振光穿过被测样品时，其中的光学活性物质会使偏振光的振动面产生旋转，旋转角的大小与样品层的厚度、入射光的波长以及温度、溶剂的pH值和光学活性物质的浓度相关。传统的偏振计一般在规定的标准条件下进行测量，也就是，被测液体样品的光程长度定为10cm，入射光用波长为5416Å的汞绿光，测量温度定为20℃。假设样品的厚度为L(dm)，溶液的浓度为C(g/100mL)，测得的旋光度为α(°)，标准条件下手性分子的比旋光度为$[\alpha]$，那么

$$C=\frac{100\alpha}{L[\alpha]} \tag{17.68}$$

式中分子的比旋光度$[\alpha]$与温度T、波长λ和溶液的pH值相关。

利用偏振测定法进行非侵入式测量（如血糖监测）时，光源发出的偏振光必须在不被完全去偏振的情况下，穿过人体组织，到达探测器。由于人体皮肤的散射系数很大，会引起光的去偏振，因此，要使偏振光穿过包括皮肤在内的厚组织层（如1cm左右厚度），且保持光的偏振信息，这是不太可能的。如果厚度小于 4mm，那么或许可行。不过，这时组织的散射作用会造成 95％以上的去偏振，偏振测量传感器必须在如此大的去偏振情况下，能够测出千分之一度数量级的旋光度。为了避免光穿过皮肤组织所产生的这种不利效应，人们选择眼球的房水作为在体检测血糖浓度的部位。房水是一种清澈的光学生物介质，并且，眼球房水的葡萄糖浓度与血糖浓度之间具有很好的相关性。兔类动物实验证明，房水葡萄糖浓度跟踪血糖浓度变化的延时很小，只有数分钟。人眼球前房的平均厚度约为 1cm。因此，假定葡萄糖对于 $\lambda=633$nm 的偏振光的比旋光度为 $45.62^{\circ}cm^2/g$，那么，在正常血糖浓度 1mg/mL 下，测得的房水的单位光程旋光度应该约为 4.562 毫度。但是，在眼球上测量也有问题，存在角膜的双折射作用和眼球运动伪迹等影响。如式（17.68）所示，由于旋光度直接与被测物体的光程长度成正比，因此，光程长度的测定很关键，或者至少对于同一个检测对象光程必须保持恒定，不随检测位点而变化。如果在眼球上测量，对于每个病人，光线在其角膜表面的入射角也必须保持相对恒定，这样，每次测量读数时，光程长度和入射角都要固定。设计仪器时，还需要考虑一般生物组织（包括眼球）的蛋白质等其他手性分子引起的旋光度变化。另外，很多生物组织还具有双折射效应，这在葡萄糖偏振仪等的设计中也必须予以考虑。

不过，在某些生物组织特性测定中，偏振光双折射引起的“延迟”效应以及组织的偏振散射是信号，而不是噪声。例如，用激光扫描偏振仪可以测量偏振光入射到视网膜神经纤维层时所产生的延迟变化。根据统计数据，在某些区域，正常眼球的这种延迟显著大于青光眼。另外，根据各种偏振光散射作用建立的图像，还可以用于区别癌细胞和正常纤维细胞。

17.4.5 微米和纳米生物传感器的应用

微米技术，特别是纳米技术的发展正在不断推进生物传感器、人造假体及其移植以及医学诊断等领域的快速发展。在医学诊断方面，微米和纳米器件已经用于芯片实验室（lab-on-a-chip），实现血液等样本的体外高效筛查分析。同时，很多研究人员和公司正在开发用于体内抗癌药物、胰岛素泵和基因治疗的纳米技术。还有人在研究用纳米材料制作的假体装置。

纳米技术的研究前沿之一是量子点（quantum dot）。量子点是能够把电子约束在很小的三维空间内的一种半导体纳米结构，在如此小的空间中，电子的量子效应（即波动性）变得很显著，超过其经典粒子效应。室温下，量子点的三维空间尺寸一般要在 20～30nm 以下。电子被束缚之后，互相之间会排斥，没有两个电子可以处于同一个量子态。于是，量子点中的电子就会形成与原子结构非常相似的壳层轨道，并且表现出原子的各种光、电、热和化学特性。在胶状绝缘层包围下，半导体纳米

粒子可以生长成为量子点。这种量子点可以沉积在某种基质表面，如分布有金属电极的半导体晶片等基质上；也可以通过各种方式结晶成为块状固体。在电刺激或者光刺激的作用下，这两种形式的量子点都会发生某些特性变化。

量子点是20世纪80年代中期在光电应用领域发展起来的纳米晶体，其在生物医学领域的应用之一是长时间全彩色监测活细胞的变化，显示出活体生物中所发生的各种复杂过程，如胚胎的发育过程等，其精细程度前所未有。如前所述，现有的荧光成像技术使用的是天然荧光分子，如水母和萤火虫上发现的有机染料和蛋白质等。这些染料的发射光波长都很宽，不同染料的光谱之间一般都会相互重叠，很难同时使用3种以上荧光染料来标定不同的生物分子。并且，这些染料的荧光发射还会随时间快速衰减。半导体纳米晶体（即量子点）可以解决这些问题。与有机荧光染料相比，量子点发射光的强度高、寿命长、发射光谱窄，而激发光谱却较宽。这样，只要用单波长的激发光源，就可以同时激发包含不同大小量子点的混合物发射出各种不同波长的荧光，用于检测和彩色成像。

以上介绍的是荧光量子点在生物医学检测中的应用，其实，这种微米和纳米粒子可以用多种材料制成，而且可以与各种不同的光学技术结合应用。例如，金或银等金属的纳米微粒可以与Raman光谱仪相结合，构成所谓的表面增强Raman光谱仪(Surface-Enhanced Raman Spectroscopy，SERS)，它比普通Raman光谱仪的灵敏度要高百万倍。这种金属微米或纳米粒子可以注入癌变组织中，成为一种吸收剂，吸收红外照射光的能量，加热肿瘤癌细胞，并将它们杀死。纳米微粒的发展才刚刚起步，这些微粒与光相结合可以形成的其他生物医学应用还有待于开发。

17.5 激光热疗的基本原理

激光热疗的原理就是将光子的能量转变为生物组织物质相的吸收能量。这种能量转变主要表现为组织温度的非均匀上升，从而导致一系列热力学过程，利用这些过程可以产生光热凝固和组织消融等治疗作用。还有一种应用是利用组织吸收的光能量激活某些内源性或外源性光敏剂，从而选择性地诱发某些光化学反应，这称为光动力疗法（PhotoDynamic Therapy，PDT）。

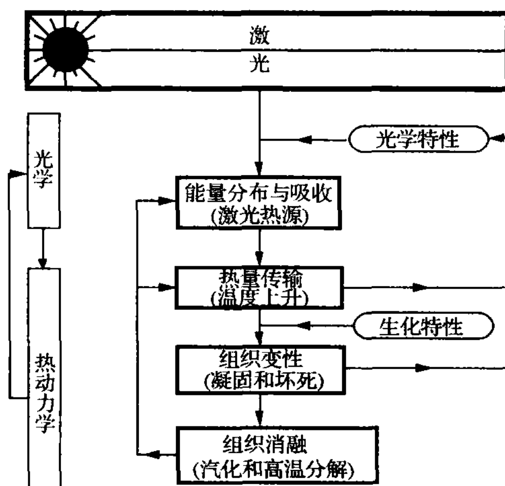


图17.13 激光与生物组织的热作用过程中光—热—化学变化的流程图

激光与生物组织之间的相互作用其实是光学过程与热力学过程的结合，图17.13概括了这些作用过程。激光照射在生物组织上时，光子进入组织，光的能量

就在组织中扩散。这种能量扩散的分布与生物组织的散射、吸收和折射等光学特性有关。其中一部分光能被组织吸收，并转变为热能，使入射激光成为一种分布式热源。此热源会引发不均匀的热传输过程，使组织升温，通过热的传导、对流和辐射等作用机制，在组织中形成与时间和空间相关的温度分布。这种温度分布的结构取决于组织的传导率、热容量、对流系数和辐射系数等热特性。

随着热量在生物组织中的不断积累和传输，组织的温度会达到某个阈值，超过这个阈值之后，就会引发不可逆的热损伤。热损伤会导致酶和蛋白质的变性，从而发生组织凝固，并最终致使细胞坏死。这种热化学损伤同时使生物组织的各种特性，尤其是光学特性，开始产生变化。组织特性的变化反过来又会影响组织的光能吸收和分布过程。

这些过程进一步发展下去，就会引发组织消融。也就是，随着局部温度的持续上升，达到另一个阈值，如果此时的热量积累速度继续大于组织传输热能的速度，那么，就会引发组织水分的猛烈汽化，并伴随大分子的高温分解，从而导致组织消融，也就是组织切除。

本节将介绍激光加热生物组织至消融的热力学过程，以及描述这个过程的各个控制方程。首先，讲述热传导方程及其一般边界条件和初始条件；然后，再介绍用于定量分析生物不可逆热损伤的 Arrhenius-Henriques 模型。

17.5.1　激光加热过程中的温度场

17.2 节已经介绍过激光能量被生物组织等介质吸收的过程。受到激光照射的物体以速率 Q_L 吸收光能之后，立即会形成向各个不同方向传播的热能流量。由于入射光束存在沿径向变化的光强分布以及轴向的光吸收，Q_L 在这两个方向上的分布都不均匀。下面推导物体物质相的能量变化率平衡方程。

假设图 17.14 所示是受到激光照射的物体中的一个无限小单元，该单元的热能存储速率 U 取决于输入热流与输出热流之差，设激光能量的吸收率为 Q_L，其他各种能量变化率之和用 Q_o 表示。单元的流入与流出热流之差的极限等于热流矢量散度的负值。于是，物体单元的能量变化率平衡方程可以写为

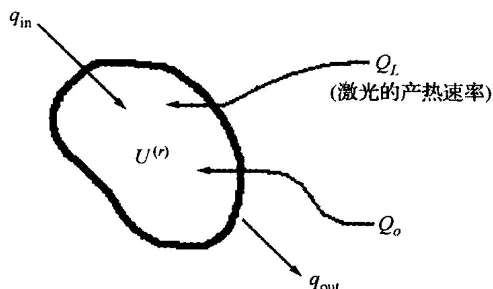


图 17.14　物质相无限小控制体积单元中的热流和能量

$$\partial U/\partial t = -\vec{\nabla} \cdot \vec{q} + Q_L + Q_o \tag{17.69}$$

式中　$\vec{\nabla} \cdot \vec{q}$——热流的变化量。

热能的状态方程为

$$U=\rho^{CT} \tag{17.70}$$

式中 T——热力学温度，单位为 K；

ρ——质量密度，单位为 kg/m^3；

C——物质的比热容，单位为 J/(kg·K)。

与光子相的推导过程一样，此处也需要一个热流的唯象关系式。常用的关系式就是 Fourier 定律，即

$$\vec{q}=-\kappa\vec{\nabla}T \tag{17.71}$$

式中 κ——物质的热传导率，单位为 W/(m·K)。

将式（17.71）和式（17.70）代入式（17.69），就得到热传导方程

$$\partial(\rho CT)/\partial t=\vec{\nabla}\cdot\kappa\vec{\nabla}T+Q_L+Q_o \tag{17.72}$$

结合适当的初始条件和边界条件就可以求解此方程，从理论上估算激光照射生物组织时所形成的温度分布。

生物组织的其他能量变化率 Q_o 可以写为

$$Q_o=Q_b+Q_m \tag{17.73}$$

式中 Q_b——血液流动带走热量的速率；

Q_m——新陈代谢产生热量的速率。

这两项热量变化此处可以忽略不计。除了长时间低功率激光治疗等应用以外，一般，这两项数值比激光热源产生的 Q_L 小得多。

很多生物组织（包括人体组织）的物质特性 κ 和 ρC 的值都已测定。通常这些参数与温度和水分有关，这是式（17.72）成为非线性方程的因素之一。还有一个非线性因素就是与温度相关的光学特性，一旦存在，它将起主导作用。

17.5.2 组织的热凝固和坏死——热损伤模型

定量描述热变性过程的模型最早出现于 20 世纪 40 年代。早期的研究应用单因素动力学速率模型，将凝固、变性和坏死等与生物组织热损伤相关的各种不可逆生化过程结合起来，粗略地用单个方程来描述。并且，定义了一个无量纲函数 Ω，作为描述组织损伤程度的特征参数，该模型假设函数 Ω 的变化率遵循 Arrhenius 关系式，即

$$\frac{d\Omega}{dt}=A\exp(-\frac{E}{RT}) \tag{17.74}$$

式中 R——普适气体常数，单位为 cal/mole；

T——热力学温度，单位为 K；

A 和 E——实验测定的常数，单位分别为 1/s 和 cal/mol。

将式（17.74）重排并积分，可以得到时间段 t 上积累的总损伤量，即

$$\Omega = A\int_0^t \exp(-\frac{E}{RT})dt \tag{17.75}$$

根据猪皮肤实验的结果，如果将 $\Omega=1.0$ 作为皮肤完全坏死，$\Omega=0.53$ 作为不可逆损伤的阈值，那么，常数 A 和 E 的值分别为 $A=3.1\times10^{98}\,s^{-1}$，$E=150000$cal/mol。

与这些值相对应的温度阈值大约为 45℃。除了猪皮肤以外，其他生物组织的有关参数值也已有报道。例如，对于人体动脉血管壁，A 和 E 的值分别为 $A=4.1\times10^{44}\,\mathrm{s}^{-1}$，$E=74000\,\mathrm{cal/mol}$。与这些值相对应的 $\Omega=1.0$ 定义为组织学分析可以测得凝固损伤时的阈值。应该注意，凝固就是胶原蛋白变性，是与皮肤烧伤不同的一种损伤。

根据热传导方程（17.72）求解得到温度场之后，再代入式（17.75），就可以估算生物组织中各点的积累损伤。所谓的温度场就是描述组织中每一点温度变化的数据。

17.5.3　光热消融

激光照射下的生物组织，其温度并不能持续地无限上升。当温度上升到某个阈值之后，如果激光产生的热积累速率仍然继续大于热传输机制可以传送的速率，那么，就会发生组织水分的剧烈汽化，结果形成空泡，随后在更高的温度下产生分解。组织剧烈汽化、空泡化和大分子高温分解的整个过程就是激光消融过程。这是激光手术切除生物组织的主要机理。

生物组织一般都含有大量水分，这说明激光消融过程的主导部分就是剧烈的汽化过程。因此，下面对高温分解只作简单介绍，主要讲述汽化过程。首先，将消融问题作为一个具有移动边界的非线性数学问题，推导消融界面上的能量平衡方程；然后，在结束时再提一下高温分解问题。

消融界面的能量平衡方程

热传导方程独自并不能用于计算组织消融前沿随时间变化的动态过程。实际上，消融前沿的移动会影响热传导过程。因此，必须建立消融前沿的能量平衡方程，才能计算其移动过程及其对于温度场的影响。

假设图 17.15 所示是激光照射下的生物组织中的某个单元，它具有单位截面积，

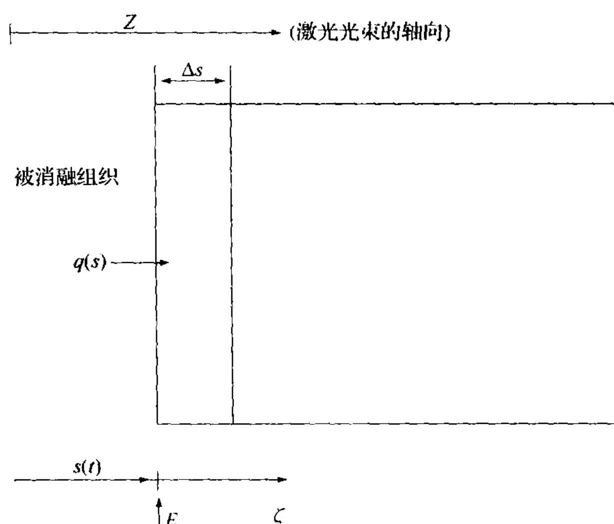


图 17.15　厚为 Δs 的小块生物组织单元消融时所需的能量（即消融界面上的能量平衡）

厚度为 Δs，并且，其温度达到了消融阈值，在接下来的 Δt 时间内就会产生汽化。那么，利用式（17.71）表示的热流 Fourier 定律，经过数学推导，可以得到如下消融边界上的能量平衡方程：

$$\kappa\frac{\partial T(s,t)}{\partial z}=\rho f_L L\frac{\mathrm{d}s}{\mathrm{d}t} \tag{17.76}$$

式中 f_L——水的含量系数。

可见，此方程是一个非线性方程，方程中的消融表面温度梯度与消融前沿位置的变化率相关，而后者事先是未知的。求解此方程需要用迭代方法，并且，要在适当的边界条件和初始条件下，同时求解热传导方程和消融前沿能量平衡方程。

式（17.76）是描述消融的热力学动态过程的基础，结合适当的边界条件和初始条件，将其与热传导方程一起求解之后，就可以估算消融前沿的位置及其移动速度。并且，还可以获得如下所述的有关生物组织内部温度分布的重要信息。

注意，在消融进行过程中，式（17.76）右边的 $\mathrm{d}s/\mathrm{d}t$ 项总是正的，这意味着方程左边的温度项 $\partial T/\partial z$ 也一定为正。因此，组织表面下方的温度一定高于表面温度。也就是，表面以下组织必定被过加热到比表面消融温度更高的温度，这样，组织表面就可以获得来自下方的热能供给，使消融过程得以进行。这与热力学第二定律的一种推论是一致的，该推论表明热只能从温度较高处传向温度较低处。实际上，高于消融阈值的温度将表面下方的组织置于一种亚平衡状态，这种状态受到内部组织状况的干扰，使得表面以下的组织开始出现成核现象并汽化，从而突破并撕裂组织。应该注意，消融过程的启动包括了许多非平衡成核过程，上述方法并没有考虑这些过程。

17.5.4 纯吸收介质中激光加热方程的解析解

17.5.3 节讲述了激光消融的控制方程。本节将这些控制方程用于一维半无限大纯吸收介质。消融前加热时期的控制方程可以写为

$$\frac{\partial\rho c T}{\partial t}=\frac{\partial^2 T}{\partial z^2}+\mu_a I\mathrm{e}^{-\mu_a z} \tag{17.77}$$

式中 μ_a——吸收系数。

在组织消融发生之前，也就是表面温度达到消融的阈值温度 T_{ab} 之前，此方程成立。这里不分析消融过程的细节。

20 世纪 70 年代末，有人求解了一个轴对称三维无量纲的热传导方程，该方程描述消融前吸收介质中高斯分布激光束的加热过程。如下这些关系式可以将控制方程转化为无量纲活动坐标系下的方程：

$$\theta=\frac{T-T_0}{T_{ab}-T_0},\ \zeta=(I_0 c/kL)z,\ \tau=(I_0 c/kL)(I_0/\rho L)t \tag{17.78}$$

式中 θ——无量纲温度；

ζ——无量纲活动坐标，其原点定为消融前沿；

τ——无量纲时间。

B 为无量纲吸收参数，λ 为无量纲加热参数，返两个参数定义为

$$B=(kL/I_0c)\alpha$$
$$\lambda=c(T_{ab}-T_0)/L \tag{17.79}$$

利用空间变量 ξ 的拉普拉斯变换，可以求得无量纲控制方程的解析解，该解如下：

$$\theta(\xi,\ \tau)=\frac{1}{B\lambda}\left\{2B\sqrt{\tau}ierfc\left[\frac{\xi}{2\sqrt{\tau}}\right]-\mathrm{e}^{-B\xi}\right.$$
$$\left.+\frac{1}{2}\mathrm{e}^{B^2\tau}\left(\mathrm{e}^{-B\xi}erfc\left[B\sqrt{\tau}-\frac{\xi}{2\sqrt{\tau}}\right]+\mathrm{e}^{B\xi}erfc\left[B\sqrt{\tau}+\frac{\xi}{2\sqrt{\tau}}\right]\right)\right\} \tag{17.80}$$

此式表明温度场是无量纲空间变量 ξ 和无量纲时间变量 τ 的函数。消融的起始对应于 $\theta=1$，在此之前，上式成立。符号 $ierfc$ 表示函数 $erfc$ 的积分，即 $ierfc(z)=\int_z^{\infty}erfc(t)\mathrm{d}t$。使用简便易学的 MATLAB 等计算机软件工具，可以方便地求解方程（17.80）。

图 17.16 显示了该方程的一组解，这些曲线表示了不同时刻下温度分布随深度变化的情况。

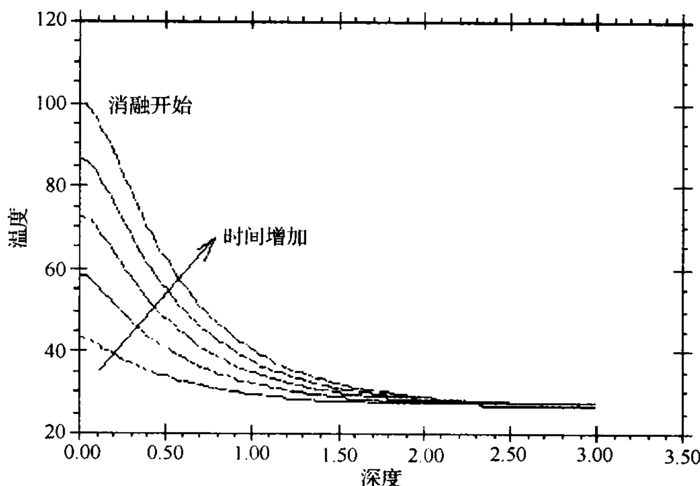


图 17.16　消融发生前各个不同时刻无量纲温度随无量纲深度变化的曲线

令式（17.78）中 $\xi=0$ 处有 $\theta=1$（即 $T=T_{ab}$），就可以得到如下超越代数方程：

$$B\lambda=\frac{2}{\sqrt{\pi}}B\sqrt{\tau_{ab}}+\mathrm{e}^{B^2\tau_{ab}}erfc[B\sqrt{\tau_{ab}}]-1 \tag{17.81}$$

用数值方法求解此方程，可以得到消融发生的起始时刻 τ_{ab}。注意，对于较大的 $B\lambda$ 值，此方程接近线性，斜率为 $\frac{2}{\sqrt{\pi}}$。

这里的分析没有考虑组织的散射效应。应用 17.2 节介绍的漫射近似法可以求得 Q_L 的解析解。但是，如果将该 Q_L 解代入式（17.72）来解析求解温度分布，那是非常麻烦的。因此，温度场的求解通常用有限差分或有限元等数值方法。

17.5.5 生化损伤的分析

前面介绍过预测组织损伤的 Arrhenius 模型，并在 17.5.2 节中定义了损伤方程。由于式（17.75）中参数 A 和 E 的值非常大，因此，数值计算时，将此式的积分写成如下等价形式比较方便：

$$\Omega(z,\ t)=\int_0^t \exp\left(\ln A-\frac{E}{RT(z,t)}\right)\mathrm{d}t \tag{17.82}$$

前面已提到过两组 A 和 E 数值，一组是猪皮肤的数据，另一组是人体血管壁的数据。如果将上一节求得的 $T(z,\ t)$ 的解析解代入上式，就可以计算 $\Omega(z,\ t)$ 的值。消融前的解只能用于消融发生之前，消融发生之后就要用消融时期的解。

将式（17.82）左边的 Ω 设为 1，并求解 z，就可以得到任意给定时刻 t 下，组织损伤的范围，也就是损伤前沿的位置。或者，对于任意给定的位点 z，按照可控的变时间步长求式（17.82）的积分，直到 $\Omega=1$ 为止。用这种方法可以计算损伤前沿位置随时间变化的曲线。在计算过程中，根据积分上限和下限值之差来控制积分时间步长的变动。例如，假设光源是持续时间为 500ms 的 5W 激光脉冲光源，分别用猪皮肤和人体血管壁的 A 和 E 数据，那么，用上述方法计算得到的消融前沿位置随时间变化的曲线如图 17.17 所示。详细内容此处不作介绍。

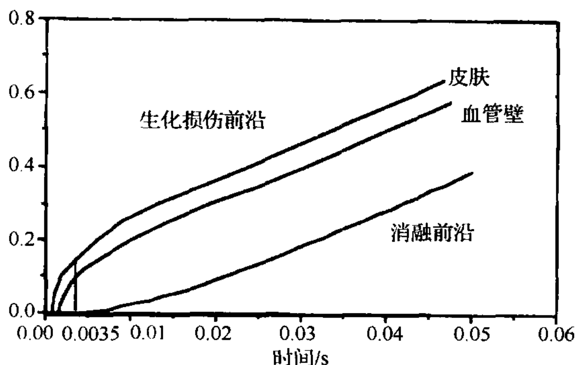


图 17.17 生化损伤前沿和消融前沿的推进过程

注：在同样的激光照射下，皮肤中消融前沿以外的热损伤厚度比血管壁中大。

图中 0.0035s 左右的垂直线表示消融起始时刻，可见在消融开始之前，组织已受到生化损伤。消融开始之后，生化损伤前沿与消融前沿之差就表示组织的实际损伤厚度。重要的是，由图中可见，根据 2 组不同的 A 和 E 数据计算得到的损伤厚度之间存在差别。显然，在同样的激光照射下，热损伤在皮肤中的推进速度比在血管壁中要快，这是因为血管壁和皮肤的损伤机制有所不同。

例 17.9 假设用温度为 T 的恒温光源加热皮肤，如果光照持续时间分别设为 $t=1\ \mu s$，$t=1ms$，$t=1s$，$t=60s$ 和 $t=30min$，那么，要在光照期间使组织达到热损伤（即 $\Omega=1$），所需的加热温度至少应该分别设为多少？也就是求临界加热温度 T_c（单位为℃）。请列表显示 T_c 随光照持续时间变化的各个数据。

解： 给定光照持续时间 $t=1\mu s$、1ms、1s、60s 和 30min，普适气体常数 $R=2cal/(gm\cdot℃)$，现在计算各个 t 值下 $\Omega=1$ 时的临界温度 T_c。

假设使用猪皮肤组织的 A 和 E 数据，分别约为 102.72 和 74000，则有

$$\Omega(z,\ t)=\int_0^t \exp(\ln A-E/(RT(z,t)))\mathrm{d}t$$

$$1=\int_0^t \exp(\ln A-E/(RT_c))\mathrm{d}t$$

$$1=\int_0^t \exp(102.72-74000/(2\times T_c))\mathrm{d}t \tag{ex. 1}$$

$$1=\exp(102.72-74000/(2\times T_c))\int_0^t \mathrm{d}t \tag{ex. 2}$$

$$1=\exp(102.72-74000/(2\times T_c))t \tag{ex. 3}$$

求解式（ex. 3），就可以得到对应于任意光照持续时间 t 的 T_c 值。部分 t 和 T_c 的对应数据如下表。

光照持续时间/s	临界温度 T_c/℃
$1e^{-6}$	143.1548
$1e^{-3}$	113.153
1	87.2
60	74.56
1800	72

■

17.5.6 汽化效应和消融温度

17.5.3 节讲述消融的能量平衡方程时曾经引入了水的含量系数 f_L。除了汽化作用以外，该系数对于表面成核现象造成的物质排出也有影响。如前所述，不经过汽化，物质块也可以通过高温分解排出。由于高温分解需要更高的消融阈值温度，消耗更多的能量，因此，高温分解是影响消融速度的另一个因素。

本章不介绍消融问题的全解，不过，通过消融速度的稳态解，我们可以了解上述现象对于消融速度的影响。稳态解的方程可以直接推导得到，为

$$v_{ss}=\frac{I}{\rho c\Delta T_{ab}+\rho f_L L} \tag{17.83}$$

给定 I、ρ、c 和 L 的值，就可以计算参数 f_L 和 ΔT_{ab} 对于 v_{ss} 的影响。例如，根据水的物理性质，I 定为 267.8W/cm^2，于是，当 $\Delta T_{ab}=100$℃且 $f_L=1$ 时，$v_{ss}=$ 1mm/s。图 17.18 显示了 f_L 从 0.01 变化到 1.0 时，消融温度对于消融速度的影响。无论是消融温度的上升还是 f_L 值的增加都会使消融速度减小，这从式（17.83）就可以看出。但是，显然 f_L 值越大，消融温度对于消融速度的影响就越小。例如，当 $f_L=0.1$ 时，如果温度从 100℃变到 500℃，消融速度就从 4mm/s 降为 1.15mm/s；然而，当 $f_L=0.8$ 时，对于同样的温度变化范围，消融速度只有 1.20～0.69mm/s 的变化。因此，如果 f_L 表示组织中的水含量，那么，只有低水分（如 30%以下）的组织才需要考虑消融温度的升高会降低消融速度。根据图 17.18 中各个不同 f_L 值对应的消融速度曲线上 $\Delta T_{ab}=200$℃处的数据，可以作出图 17.19 所

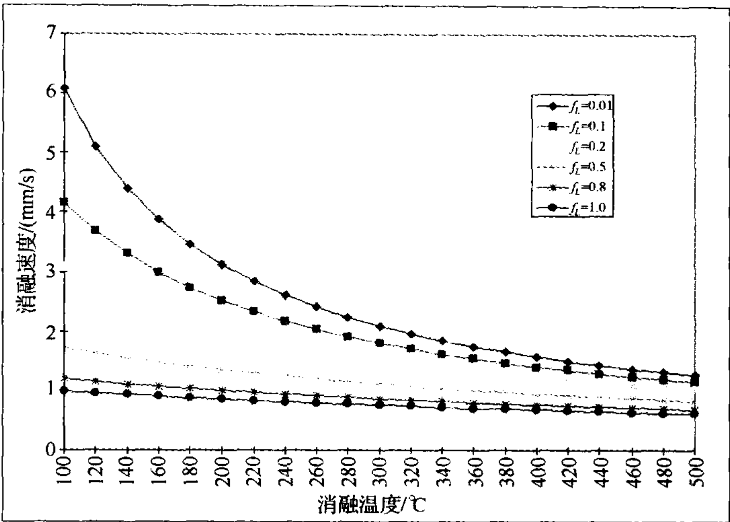


图 17.18 各种水分不同的组织中消融温度对于消融速度的影响

示的消融速度随 f_L 变化的曲线，图中的 f_L 取值比图 17.18 多。

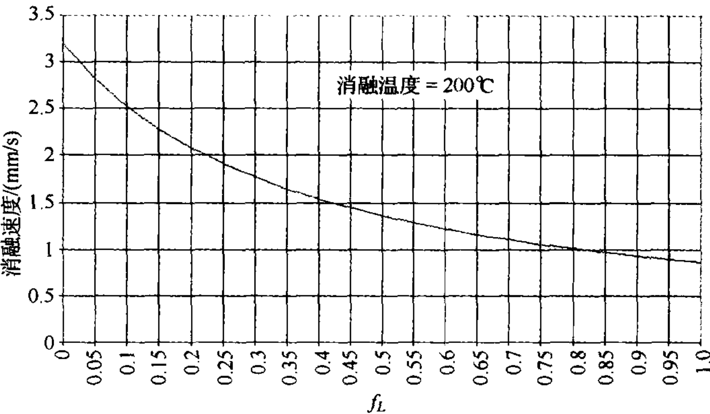


图 17.19 水分含量系数 f_L 对于消融速度的影响

由图 17.19 可知，$\Delta T_{ab}=200$℃时，当 f_L 从 0 变化到 1 时，消融速度则从 3.2mm/s变化到 0.86mm/s；也就是，消融速度可以超过其在 $f_L=1.0$ 处值的 3 倍。图 17.18 的各条曲线在较高的消融温度处，相互之间靠得比较近，这表明消融温度越高，f_L 值的变化对于消融速度的影响就越小。

17.6　光纤和波导在医学中的应用

早在耶稣基督诞生之前数百年，人类就已经使用各种刚性导管检查人体内腔。但是，直到 19 世纪，人们才在导管内加上了照明，也就是用 45°反射镜将烛光投射到导管内。20 世纪初出现用于图像传输的多透镜之后，产生了可以插入人体的半柔性导管。20 世纪 20 年代末又出现了用透明细玻璃丝制作的光纤探头，但这种探头被搁置了 20 多年，在 50 年代才开始受到广泛关注。1959 年有人研制出了第一台基于光纤技术的医学仪器，并用于临床病人的诊断，这就是柔性光纤胃窥镜。20 世纪 60 年代又诞生了激光器，70 年代初光纤在通信领域的应用迅速发展。所有这些进展推动了如今在医学领域广泛使用的光纤探头和内窥镜技术的发展。

17.6.1　光纤和波导的基本原理

17.2.1 节讲述了光在介质中的几何传播。当时，我们将光看做光线，推导了 Fresnel 方程，用于描述光束在两种物质交界面上的相互作用。结果表明，随着两种物质折射率的变化以及光线入射角的变化，交界面上反射光和折射光的量也会发生变化。虽然本节将介绍的光纤中光的传输问题很复杂，不过，这种传输也可以用简单的几何模型来解释。如图 17.20 所示，将光纤看做一根透明的细丝，假设光纤纤芯的折射率为 n_1，其圆柱形外壳材料（即包层）的折射率为 n_2，且 $n_1>n_2$，那么，根据 Snell 折射定理，当光线在纤芯与包层交界面处的入射角大于某个临界角时，光在此界面上就会发生全内反射，这样，光线就会沿着纤芯传播。为了使光线在纤芯与包层界面处能够发生全反射，入射光在光纤端面相对于芯轴的入射角，必须在某个锥度范围之内，该锥度角的最大值 θ_0 定义为

$$n_0\sin(\theta_0)=(n_1{}^2-n_2{}^2)^{1/2}=NA \tag{17.84}$$

此式也是数值孔径 NA 的定义。硅光纤的 NA 通常为 0.2～0.4。由此式可见，增加纤芯折射率与包层折射率之间的差别，可以增大 NA 值，也就可以增大光纤接收入射光的锥度角。NA 和接收锥度角越大，光纤能够传输的具有不同反射角的光线就越多，也就是传输模式越多，每种传输模式具有不同的光束折线路径。如果纤芯直径很小，那么，光纤就只能产生单一的传输模式。

为了满足不同的应用需求，光纤产品具有各种不同的结构、形状和材质。例如，用于前列腺组织凝固等治疗的光纤，其尖端被加工成可以产生侧射（side-firing）光束。光纤尖端还可以做成逐渐缩小的锥形，提供点光束；也可以做成发散的尖端，提供均匀的大面积光照。一般，按照纤芯折射率分布的不同，以及单模式还是多模式传输，可以将光纤分成各种不同的类型。例如，如图 17.21 所示，如果整条光纤纤芯的折射率为恒定不变的常数，那么这种光纤称为阶跃型光纤（step-index fiber）；如果光纤的折射率不是常数，而是从纤芯中心到包层界面呈抛物线分布，中心折射

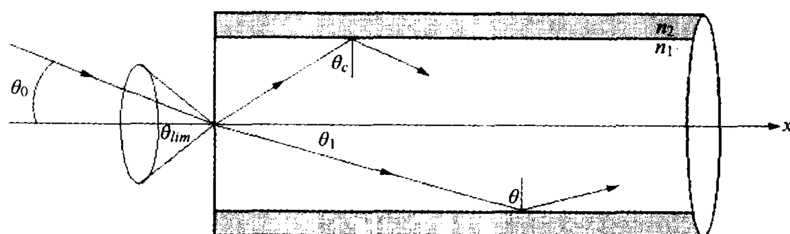


图 17.20 光纤的基本传输模式（其中圆柱形纤芯的折射率为 n_1，包层的折射率为 n_2，且 $n_1>n_2$）

率高，边缘折射率低，那么这种光纤称为梯度型光纤（graded-index fiber）；如果光纤纤芯很细，NA 值很小，那么就是单模光纤（single-mode）。与粗细相当的阶跃型光纤相比，梯度型光纤可以将模式色散减小到 1%，并增加带宽。这是常用光纤的分类法，除此之外，有时还用到某些特殊结构的光纤，例如，下面将介绍的用于内窥镜成像的相干光纤束、用于检测和诊断的单光纤或非相干光纤束以及用于治疗的高功率光纤等。

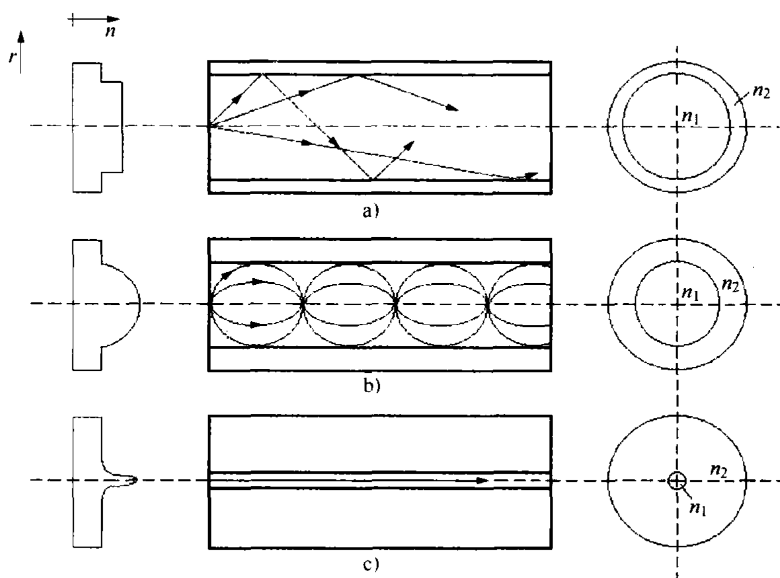


图 17.21 几种主要光纤类型
a）阶跃型光纤 b）梯度型光纤 c）单模光纤

关于光纤材料，需要考虑波长、功率和生物相容性这 3 方面问题。治疗时，无论是使用连续激光还是脉冲激光，都要考虑光纤对于激光功率的影响。光纤中含有的杂质会吸收光能，一方面减少到达探头出口的检测光的能量，另一方面还会加热光纤，致使光纤受损。除了杂质以外，光纤材料本身可能只适用于某些波长的光，不能用于其他波长的光。例如，标准硅（即玻璃纤维）可以很好地传输可见光波长

范围内的光，但却会吸收波长大于 2.0μm 的光，因此，不能用于红外线的传输。其他光纤，如锗、蓝宝石、氟化钡和空芯波导（如镀金光纤）等，则在可见光和红外区域都可以使用。但是，如果用于人体，除了上述问题以外，还需要考虑光纤的许多其他问题，如不能太硬，不能有较强的吸湿性（即溶于水），更不能有毒。因此，光纤材料的生物相容性很重要，必须对人体无毒，在体内必须能够正常工作，不能被溶解或者断裂。

17.6.2　光纤传像束

将很多根光纤的两端按照一定的规则排列起来，组成一束，就可以做成光纤传像束。其中，每根光纤在两个端头的位置都必须严格一一对应，而中间部分则可以松散，这样就构成了既柔软又可以任意弯曲的有序排列的传像束。如果将图像投射到传像束的一端，那么，在其另一端就可以得到同样的图像。光纤传像束是内窥镜的关键部件。

内窥镜可以通过尽可能小的创口，将体内器官的图像传输到体外。使用时，用一束普通的无序排列光纤束将光源导入，照亮体内需成像的部分，同时用一束光纤传像束将图像传输出来。一般用白光光源照明，这样就可以得到具有真彩色的人体组织图像。

采光量和图像分辨率是决定传像束图像传输质量的两个主要因素。每根光纤能够采集到的光的功率大小由纤芯的直径和数值孔径 NA 决定。纤芯直径和 NA 值较大，采集到的光强就大，医生就能够较好地观察到体内被照明物体的图像。图像分辨率代表所呈现的物体的细节程度，它由每根光纤纤芯的直径 d 决定。传像束能够分辨的最小细节约等于 $d/2$。分辨率定义为每毫米能够分辨出来的线条数目。因此，纤芯的直径越小，分辨率就越高。由于受到分辨率的限制，物体上的直线在图像中可能呈现为 Z 字形折线。

传像束中的每根光纤都必须有包层，用于尽可能减小和消除光纤之间的串扰。早期的内窥镜采用没有包层的光纤，成像质量就很差。无包层光纤中传输的光线会从纤芯中逸出，并进入其他光纤，这种串扰会引起图像不同部分的重叠，使图像模糊不清。

采光量大和图像分辨率高这两个要求之间存在矛盾。纤芯直径越大，传输的光就多，但分辨率却越低。另外，如果涂上较厚的包层，虽然可以避免不同光纤之间的串扰，但却限制了采光量和图像分辨率。因此，实际应用中要权衡利弊。常用的光纤纤芯直径是 10～20μm，包层厚为 1.5～2.5 μm 数量级。

其他一些因素也会影响光纤传像束的成像质量，例如：某些偏离的光线会穿过包层进入其他光纤的纤芯，这些光线叠加起来形成了背景噪声，从而降低了图像的对比度。传像束中有缺陷光纤的存在会产生更严重的问题，如果某根光纤坏了，不能传输光线，那么图像中就会出现一个固定不变的暗点。另外，如果各个光纤在两端的排列不是严格对应，那么，图像就会变形。

光纤传像束的两端都可以加上透镜，用于调节放大倍数。虽然原理上可以通过调节物镜与光纤束之间的距离来改变传像束的聚焦，但是，为了简便，通常使用景深较大的固定焦距。另外，显示在监视器上的图像可以用 VCR 实时记录下来。

如今市场上已经有很多种专用光纤内窥镜产品，可以用于各种不同人体器官的成像，例如：静脉和动脉血管镜、关节内窥镜、支气管内窥镜、胆管镜，结肠镜、阴道镜、膀胱镜、食管镜、肠胃镜、腹腔镜、喉镜和脑室镜等。

17.6.3 光纤检测探头

在人体微创检测领域，光纤探头是 20 世纪 80 年代之后才发展起来的一种新兴技术。为了了解这种探头的应用前景，我们先来回顾一下传感器的定义及其性能要求。传感器是将某种输入参量（即被测量）转化成另一种参量（即所谓的信号）的器件。例如，通过光纤探头尖端上薄膜的位移，可以将压力信号转化为光强信号，而光强信号又可以由光探测器变成电压信号。衡量传感器性能的指标有：特异性，也就是在不受其他参量干扰的情况下，检测出某种特定参量的能力；灵敏度，也就是测出给定被测量微小变化的能力；准确度，也就是测量值与真实值之间的接近程度；另外，还有成本要低。与一般传感器相同，光纤传感器的应用也需要在一定限制条件下权衡利弊。例如，在合理的成本下可以得到 95%的准确度，但要达到 99%的准确度，就像军用技术规格所要求的那样，那么成本就可能大大增加。光纤探头有很多优点，如：体积小，速度快（因为光速很快），可见光和近红外波长的探头生物相容性好，并且安全（人体不需要电气连接）。

使用各种光纤探头时，都要先将光传入体内．光与体液和生物组织或者体内压力、体温等产生直接或间接的相互作用，从而引起光束的改变，变化了的光束再反传回检测系统。输入光束和输出光束可以通过各自独立的光纤传输。如果输入和输出光束的波长不同，也可以用同一根光纤传输，荧光光纤探头就是如此。光纤传感器可以用于 17.3 节和 17.4 节所介绍的各种化学和物理参量的测量。不过，这里我们不逐一介绍各种光纤探头的具体应用例子，而是将光纤探头分为直接和间接两大类来介绍。

直接光纤传感器中，光直接与被测样本产生作用。例如，理论上，通过测量样本的光吸收，可以检测血氧含量等参量，定量分析血样。图 17.22a 所示的相距间隔固定的两根侧射光纤，或者图 17.22b 所示的渐逝波光纤都可以用于这类检测。渐逝波光纤的部分包层被剥去，这样，沿着纤芯传导的部分光线就会跑出光纤，与样本发生作用，然后再返回光纤的纤芯（见第 9 章）。图 17.22c 所示是另一种直接测量方法，它利用生物组织的自发荧光（即 Raman 光谱）进行检测。其中的光纤束同时将输入光传送给组织，并将组织发射的荧光接收回来，用于早期癌症的癌变组织诊断。17.3.3 节介绍的 Fabry-Perot 干涉仪等各种干涉仪也可以通过半反射镜与光纤相结合，这样，由体内温度或者压力等变化引起的各种光纤参量的物理变化，就

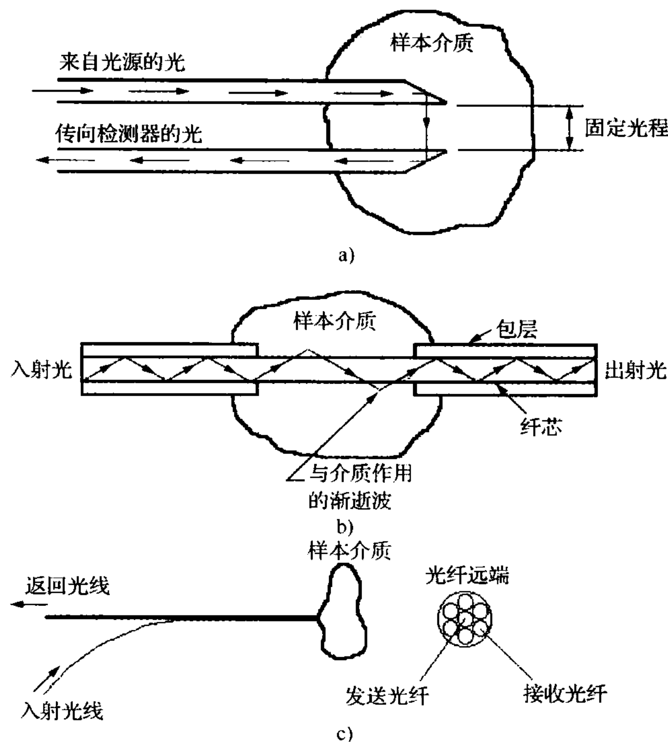


图 17.22　几种直接测量型光纤探头

a）使用侧射光的吸收型探头　b）渐逝波探头，其中的光线从纤芯传到样本，再从样本传回纤芯　c）直接测量生物组织自发荧光（即 Raman 光谱）的多光纤探头，用于鉴别正常组织和癌变组织

可以通过光干涉条纹的变化直接测量出来。最后，17.2.3 节所述的利用光纤探头测量生物组织光学特性的方法也是一种直接测量样本反射光的方法。注意，这里只是简单介绍各种光纤探头的原理，实际应用中，需要在人体噪声环境下检测各种微弱信号，必须经过大量设计工作，才能使光纤探头达到所需的特异性、准确度和灵敏度。

间接光纤检测通过安装在光纤远端的一种微型传感器进行测量，这种光纤有时被称为光极（optode)。光极中的光与微传感器产生相互作用，而微传感器再与样本相互作用。例如，图 17.23a 所示是一种位移光极，其尖端薄膜产生移动时，通过反射进入光纤的光就会减少，于是，这种光极就可以用于监测压力或者温度的变化。另外，如图 17.23b 所示，还可以在光纤远端的薄膜内加入荧光物质，其产生的荧光强度就可以反映血液中某种特定物质的浓度。这里也只是简单讲述这些间接测量方法的原理，其实，达到一定精确和灵敏度要求的实际产品是相当复杂的。总之，间接光纤测量方法的特异性通常比直接测量方法要高，但其代价是探头比较复杂。

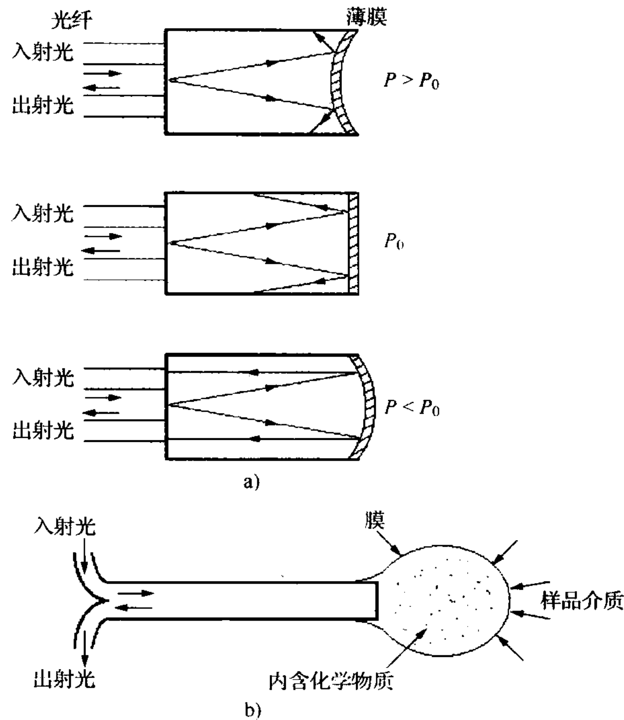


图 17.23 间接测量型光纤探头

a）用于压力或者温度测量的反射薄膜 b）光纤远端薄膜内包含某种特殊化学物质的光极，它利用对于待测物具有特异性的荧光物质来测定待测物的浓度

17.7 生物医学中的光学成像

近一个世纪以来，医学成像的发展引起了医学领域的彻底变革，使医生能够在无创的情况下清楚地“看见”人体内部的病变并迅速做出诊断，对于疾病的治疗效果具有深刻的影响。

医学成像起源于 1895 年德国物理学家 W. C. Roentgen 偶然发现 X 射线的事件。由于这项重要发现，Roentgen 在 1901 年获得了首届物理学诺贝尔奖（http://www.almaz.com/nobel/）。现在我们知道，直接利用 X 射线投影就可以容易地呈现出人体的解剖结构，如今仍在使用的 X 射线胸透成像就是一个例子。20 世纪 70 年代计算机断层扫描（CT）成像技术的发明标志着现代化医学成像技术的开始。当时，英国人 G. N. Hounsfield 基于美国人 A. M. Cormack 创立的理论，试制了第一台计算机图像重建系统。他们两人于 1979 年同时获得了诺贝尔医学奖。发明 CT 的依据是：根据各个不同角度所观察到的物体的信息，可以计算出其截面的图像，这就是图像“重建”原理。

CT 的出现激发了其他各种断层成像新技术，甚至三维成像技术的发展。例如，图像重建技术用于传统核医学成像就产生了正电子发射断层成像（Positron Emission Tomography，PET）和单光子发射计算机断层成像（Single Photon Emission Computed Tomography，SPECT）；而用于核磁共振则产生了磁共振成像（MRI）。

各种不同的成像技术可用于检测生物组织的不同特性。例如，X 射线成像主要检测电子密度；MRI 检测的是质子密度及其相关的弛豫特性；超声成像检测声阻抗；核成像检测服用放射性药物之后人体发出的核辐射，并由此反映体内某种特定的生化功能。各种成像技术都具有各自的优缺点，下面以乳腺癌检查为例加以说明。乳腺癌是一种最常见的恶性肿瘤，居美国妇女癌症死亡病因之首。目前还没有发明防止乳腺癌的方法，早期诊断和治疗是提高其治愈率的最好方法。现在，临床上常用乳房 X 线摄影和超声成像联合检测乳腺癌。乳房 X 线摄影是目前惟一能够检测出临床上未能触及的乳腺癌的可靠方法。作为辅助工具，超声检查则用于评价 X 线摄影检测出来的肿块影或者触摸发现的肿块的囊实性，这在乳房 X 线摄影中很难判断。另外，乳房 X 线摄影具有电离辐射，且很难显示致密型乳房中的肿块。虽然多数研究结果声称超声成像比低性能的乳房 X 线摄影要有用，不过，高性能乳房 X 线摄影能够显示的许多未能触及的乳腺癌，超声成像并不能检测出来。

其他乳腺癌成像技术也正在开发。其中，MRI 用于致密型乳房的成像具有很好的前景。乳房 MRI 在区别实性和囊性病变上优于乳房 X 线摄影；在显示不同软组织结构上则与乳房 X 线摄影相当。虽然乳腺癌组织与腺体组织具有相似的磁共振特性，但是，通过静脉注射造影剂，就可以提高 MRI 的癌症检测能力。不过，乳房 MRI 费用很高，其空间分辨率低于乳房 X 线摄影，并且不能显示微钙化组织。也有人使用乳房 CT 扫描来区别良性和恶性肿瘤。但是，乳房 CT 需要静脉注射碘造影剂，且空间分辨率较低，费用较高；因此，它不适用于常规乳腺癌筛查。

17.7.1　光学断层成像

虽然光的透射成像早在 1929 年就开始应用了，但是，非电离辐射的光学断层成像却是一个新兴研究领域。虽然正常生物组织本身的光学特性就存在很大变异，不过，正常组织与病变组织的光学特性还是不同的，可以通过光学特性的检测来诊断某些乳腺癌等病变。

由于癌变组织的细胞结构和亚细胞结构存在明显变化，并且，引起弹性散射的细胞成分的大小通常相当于可见光到近红外线波长的数量级，因此，癌变组织显然会表现出光学特性的变化。例如，某些肿瘤会产生血管增生，其中的血液会增加光吸收量。利用光学造影剂还可以提高正常组织与异常组织之间的成像对比度。由于光学信息取决于生物组织的分子构象，因此，光学成像应该能够灵敏地检测和监视早期癌变。

生物组织是具有高散射性的混浊光学介质，频繁的散射会使组织中的光迅速发散。这种强散射性是生物组织光学检测难以解决的一个问题。一般，生物组织中，

可见光的散射系数约为 $100cm^{-1}$，用于医学诊断的 X 射线的散射系数约为 $0.2cm^{-1}$。透过生物组织的光可以分为 3 类：弹道光（即直行光）、蛇行光（即准弹道光）和漫射光。弹道光没有经历任何散射，直接穿过组织。就像 X 射线透射成像一样，弹道光中包含直接的成像信息。蛇行光经历少数散射，包含部分成像信息。而经历多重散射的漫射光则几乎不含直接的成像信息，反而会干扰弹道光和蛇行光。

早期光子成像（early-photon imaging）是一种光学断层成像技术。如果将漫射光挡去，只采集弹道光和蛇行光，就可以像 X 射线透视一样检测深藏在介质内的物体。这种技术使用脉冲宽度小于 1ps 的短脉冲激光照射生物组织。只有最先射出组织的那部分透射光才被允许通过光检测器，随后的透射光用快速时间选通门挡掉。由于弹道光和蛇行光通过生物组织时的光程总是比漫射光短，因此，其光子比漫射光子先到达检测器。只检测弹道光的技术称为弹道光成像。研究表明，这种成像技术只能用于厚度小于 1.4mm（即平均自由程为 42mfp）的生物组织。文献中报道的一般弹道成像技术只达到约 30mfp。因此，这种方法只适用于薄组织样本。对于厚组织，由于组织的强散射效应，弹道光强度和分辨率的损失都很大。

例 17.10 请计算弹道光穿过厚度为 30mfp 生物组织之后的衰减。如果组织的散射系数为 $100cm^{-1}$，请计算相应的厚度。

解：根据比尔定律，衰减等于 $\exp(-30)=9.4\times10^{-14}$；

组织的厚度为 $30/100cm^{-1}=0.3cm$。 ■

临床检测的乳腺组织厚约 5～10cm，必须用散射光成像的方法。研究结果表明，对于吸收系数为 $0.1cm^{-1}$ 且约化散射系数为 $10cm^{-1}$ 的 5cm 厚乳腺组织，检测器必须从至少经历了 1100 次散射的透射光中采集到足够强的信号，再用于成像。此时透射光中的弹道光和蛇行光都几乎荡然无存，因此，不能用这两种光成像。但是，如果将 10mW 的可见光或近红外激光照射在 5cm 厚的乳腺组织的一侧，根据漫射理论的计算结果，漫射透射率约为 $10nW/cm^2$（即 10^{10} 光子/$(s\cdot cm^2)$）数量级，这种光强可以用光电倍增管检测，光电倍增管能够实现单光子计数。同理，通过 10cm 厚乳房组织的漫射透射率约为 $1pW/cm^2$（即 10^6 光子/$(s\cdot cm^2)$）数量级。虽然散射系数较高，但由于吸收系数较低，因此，光的透射率还是比较大。

单纯激光成像的分辨率随着组织厚度的增加线性下降。散射光的时间分布可以用超高速扫描摄像机检测。将该分布的初期部分积分，就可以建立混浊介质中隐藏物体的图像。这种时域成像方法需要昂贵的短脉冲激光器和快速光检测器。

光学相干断层扫描成像技术（Optical - Coherence Tomography，OCT）的出现为临床检测提供了一种很有用的工具。这种成像仪实际上是一种使用短相干长度光源的 Michelson 干涉仪（见图 17.9）。干涉仪的一路光束射向被测样品，另一路光束则射向参考反射镜，光电检测器检测两路光束反射回来的光强。只有当样品臂和参考臂之间的光程差在光源相干长度范围之内时，两束反射光才会产生干涉，这样，就可以检测来自样品不同深度的反射光。OCT 系统采用外差法进行检测，也就是，直接利用多普勒频移效应，通过高速匀速扫描的方法改变参考反射镜的位置，使参考

光路的长度产生微小变化，记录干涉信号幅度随着参考反射镜位置变化的数据，此信号就反映了样品的反射强度，从而可以产生类似于 A 型超声的扫描图像。OCT 的分辨率可以达到 10μm，但穿透深度小于 1mm。

最近，OCT 已经扩展到应用多普勒频移效应进行皮肤表层血管的血流成像。血液的流动会引起入射光频率的多普勒频移，将光学干涉信号进行快速傅里叶变换，就可以得到多普勒频移数据，然后，再用此频移数据计算血液的流速。

有一种光学成像方法可用于厚组织，它就是基于光子密度波（photon-density waves）的频域法。光子密度波的控制方程是如下漫射方程：

$$\partial\Phi(r,\ t)/c\partial t+\mu_a\Phi(r,\ t)-D\nabla^2\Phi(r,\ t)=S(r,\ t) \tag{17.85}$$

式中　$\Phi(r,\ t)$ ——t 时刻在位点 r 处的光能流率，单位为 W/cm^2；

c——介质中的光速，单位为 cm/s；

μ_a——吸收系数，单位为 cm^{-1}；

$S(r,\ t)$——光源强度，单位为 W/cm^{-3}。

频域成像技术需要使用复杂的图像重构反变换算法。检测时，从多个不同位置，用调幅激光（约 100MHz）照射生物组织。并在各个照射位置下，用光检测器在组织周围多个不同位置上测量漫射光的幅度和相位。如果组织的光学特性已知，那么，用式（17.85）的漫射方程就可以估算漫射光的测量值。反过来，也可以通过测得的漫射光数据来计算组织的光学特性，这就是图像重建。图 17.24 所示的样本图像就是根据理论计算生成的漫射光数据重建的。

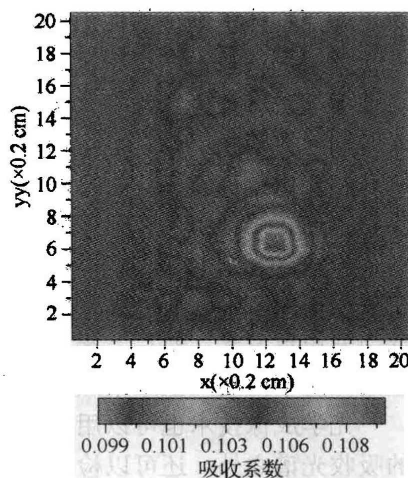


图 17.24　利用光子密度波数据重建的样本图像（在图像右下方可以看到吸收系数升高的肿瘤组织）

17.7.2　光声混合成像技术

目前正在开发的几种新型成像技术将穿透力强的声波与散射光结合起来，产生了所谓的光声混合成像技术。下面简单介绍这类成像技术中的超声调制光学层析成像、光声层析成像和声致发光（Sono Luminescence，SL）成像方法。

超声调制光学层析成像技术，利用聚焦在散射介质中的超声波，来调制穿过介质的激光。这些调制激光的特性与介质中超声调制区域的力学特性和光学特性相关。用光电倍增管采集穿过介质的调制激光，如果介质中隐含的物体的特性与散射背景介质不同，那么，通过聚焦超声波的扫描就可以得到该物体的图像。

将超声调制光学层析成像（即声光成像）反过来就变成了光声层析成像。光声层析成像技术，利用短脉冲光束照射散射介质，光束在介质中漫射并被部分吸收。如果介质中存在强吸光物体，如肿瘤，那么，该物体吸收的光会比周围区域多。被

吸收的光能转化为热能，吸收物受热膨胀会产生超声波。生成的热量越多，产生的超声波就越强。因此，较强的吸光物体会发出较强的超声波。如果用多个超声波换能器检测介质周围的声波信号，根据各个声波信号的时间分布数据就可以确定吸光物体的位置，并由此建立介质的图像。最近，这项技术已经用于小动物的脑功能成像。

物体在超声波作用下产生光的现象被称为声致发光，最早发现于 1934 年，当时是一种多泡声致发光（Multiple - Bubble SonoLuminescence，MBSL）。1990 年单泡声致发光（Single - Bubble SonoLuminescence，SBSL）发明以后，声致发光现象受到了广泛关注。虽然 SL 的机理还没有完全搞清楚，但是，人们已经知道，在超声波的作用下，当物体中细小的囊泡破裂时会发出光。小泡的初始半径约为数微米，由于声波正弦周期的负半周声压减小，小泡会膨胀到近 50μm。到了声波的正半周，声压又迅速增加。这种压差变动会使小泡迅速破裂，并伴随着发光。根据测量数据，SL 闪光时间只有数十皮秒。SBSL 非常亮，甚至在明亮的室内肉眼都可以看到，而 MBSL 则在暗室内可以看见。研究人员正在设想利用 SL 的超快闪光开发其在声聚变、声化学以及构建超快激光方面的应用。

声致发光层析成像（SonoLuminescent Tomography，SLT）是 SL 的一个应用，主要用于强散射介质的无创断层成像。它利用外部超声波扫描被测介质，在介质内部引发声致发光，然后，根据测得的发光强度数据，建立散射介质的图像。图像的空间分辨率由超声波聚集点的大小决定，缩小聚焦点可以提高分辨率。

光学成像技术也可以用于无创检测生物组织的光谱。除了其他物体测量中常用的吸收光谱之外，还可以检测散射光谱。光谱可以用于定量检测血氧饱和度等重要的生理参数。

总之，光学成像等生物医学光学技术通常具有以下优点：属于非电离辐射；能够检测功能性生理参数；检测生物组织病变的灵敏度较高；成本较低。不过，生物医学光学是一个充满挑战的新兴研究领域，需要拥有各种才能的医生、科学家和工程师们的共同努力，才能推进其发展。

习　题

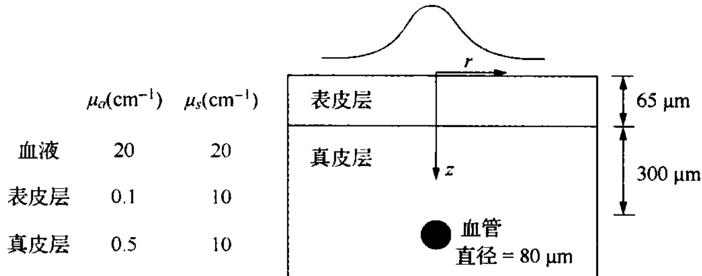
1.（a）为什么生物医学应用中需要光学传感器或光纤传感器？请至少说出 3 点理由。（b）请说出 3 个光学或光纤生物医学传感器的潜在缺陷。

2. 请说明光的偏振、冷发光、光的吸收、光的散射和 Raman 散射之间的异同。

3. 假设清澈的水组织（$n_t=1.33$）与玻璃（$n_g=1.5$）之间形成交界面，一束入射光以 45°入射角从水中射向界面，请计算此光线的出射角。如果反过来，光线从玻璃一侧入射到界面上，请计算水组织中的出射角，并解释此结果是否合理。

4. 请用 MATLAB 等软件，分别作图显示玻璃（$n_i=1.5$）与空气（$n_t=1.0$）交界面上，$R_\perp=r_\perp^2$ 和 $R_p=r_p^2$ 随入射角 θ_i 变化的曲线，并解释其含义。

5. 假设用波长为 1064nm 的高斯分布 Nd：YAG 激光束治疗血管扩张形成的鲜红斑痣（即葡萄酒色痣）。设均匀介质中光分布的模型为 $I(r, z)=I_0\exp(-\mu z)\exp(-2(r/\omega_b)^2)$，其中 $\mu=\mu_a+\mu_s$，μ_a 和 μ_s 分别是吸收系数和散射系数，并且 $I_0=1$。请根据下图所示该波长下各层皮肤组织和血管的吸收系数和散射系数，求出组织中 $r=0$ 处光强 I 和激光产热速率 Q 的表达式和曲线，即 $I(r=0, z)$ 和 $Q(r=0, z)$。并说明血液层的选择性光热解效应是如何形成的。



6. 假设某种混浊生物组织的吸收系数为 $\mu_a=0.1\text{cm}^{-1}$。请应用漫射近似法，在同一张图上做出对应于 4 个不同约化散射系数 $\mu_s=1$、10、50、100cm^{-1} 的相对光强随半径 r 变化的对数曲线，其中 r 的取值范围为 0.1～2.0cm。并说明散射效应对于光强分布的影响。

7. 请编写 C/C++或其他计算机语言程序，完成斜入射漫反射 $R(x)$ 的漫射理论计算。

8. 假设肝脏组织受到 Ho：YAG 激光脉冲的照射，请计算热传导可以忽略不计的光照持续时间。设 Ho：YAG 激光波长处的吸收系数为 30cm^{-1}，并且肝脏组织的热扩散与水相同。如果消融的阈值温度是 150℃，请计算消融的阈值通量，单位为 J/cm^2。

9. 假设用某个氦氖激光的光纤多普勒探头测量血流速度，探头与血管呈 60°角，得到的频移为 63kHz，那么，血流速度是多少？从生理学上分析，与人体内大小血管的平均血流速度相比较，该计算结果是否合理？

10. 假设你要用 Michelson 干涉仪或者 Fabry-Perot 干涉仪测量一块置于空气（$n_i=1.0$）中的生物组织（$n_t=1.33$）的厚度，所用光源波长为 780nm，并且测量仪器可以读出 20 条干涉条纹。方法是：用大功率脉冲激光切割这块组织（即消融），每个激光脉冲大约能去除 2 μm 厚的组织。当组织被完全切除时，测得其厚度是 20 μm。(a) 请通过计算说明，在不改动仪器的情况下，两种干涉仪是否都具有足够长的动态范围用于这块组织的测量。并解释所得到的结果。(b) 假设需要将这两台仪器之一用作生物组织激光切除的反馈检测仪器。为了能够测得组织切除过程中的最小厚度和最大厚度，每种仪器的干涉条纹数的范围（即最小和最大条纹数）应该为多少？为什么。

11. 在一般情况下，什么时候 Fabry-Perot 干涉仪比标准 Michelson 干涉仪更具有

优势？你可以制作一台比 Michelson 干涉仪性能差的 Fabry-Perot 干涉仪吗？为什么。

12. 什么是 Raman 光谱仪的 Stokes 频带和反 Stokes 频带？通常哪种频带可以用于检测？将 Raman 光谱仪用于生物医学检测时，需要克服的主要问题是什么？

13. 假设你想用无创测量眼睛房水葡萄糖含量的方法，来定量检测血糖浓度。如果 656nm 波长处葡萄糖的比旋光度为 41.89°/(dm・g/mL)，并且眼睛房水的光程长度约为 0.8cm，请计算 15 毫度旋光度所对应的葡萄糖浓度。从生理学上看，所得的计算值合理吗？被测病人属于正常还是患有糖尿病？

14. 在 600～1100nm 的近红外光谱区域存在氧合血红蛋白和还原血红蛋白的吸收峰。这里，暂时假设该光谱区域以光的吸收为主（注意：一般不会如此）。已知：组织样本的厚度为 2cm。在 758nm、798nm 和 898nm3 个波长下测得的氧合血红蛋白消光系数分别为 1.612、2.072 和 2.763$mM^{-1}cm^{-1}$，还原血红蛋白的消光系数分别为 3.914、2.072 和 2.072$mM^{-1}cm^{-1}$。这 3 个波长下组织的总吸收系数可以用时间分辨系统测得，分别为 0.293cm^{-1}、0.1704cm^{-1} 和 0.1903cm^{-1}。请计算氧合血红蛋白和还原血红蛋白的浓度。从生理学上看，该结果合理吗？为什么。

15. 应用本章给出的热传导方程的解析解，并利用习题 8 的参数，作出 $z=0$ 处温度随时间 t 变化的曲线，t 的取值范围为 0～10ms。且在同一张图上作出温度随时间线性变化的直线（忽略热传导效应）。并结合习题 8 计算得到的扩散时间解释图中显示的结果。

16. 假设皮肤组织被线性加热，就像脉冲激光凝固中 $T=T_0+mt$ 的情况，其中 T_0 是初始温度（如 37℃），m 是加热速率（如 10^3℃/s）。请说明如何计算热损伤的临界温度。假设光照时间相同，你觉得此临界温度应该低于还是高于恒温加热时的临界温度？

17. 为了预测组织消融的发生时刻，请作出 $B\sqrt{\tau_{ab}}$ 随 $B\lambda$ 变化的曲线，并测出"大" $B\lambda$ 值下的斜率，应该为 $\frac{2}{\sqrt{\pi}}$。假设组织的吸收系数为 100cm^{-1}，激光强度为 1000W/cm^2，请使用水的热特性计算消融起始时刻（设消融始于 150℃）。

18. 在可见光区域，生物组织的散射效应超过吸收效应。假设有一块 5cm 厚的生物组织，其散射系数为 100cm^{-1}，所用的激光波长为 0.5μm。为了能够平均检测到一个穿过组织的弹道光光子，请问入射光需要具有多少焦耳的能量？并将此入射光能量与地球的静止能量进行比较。设地球的质量为 6×10^{24}kg，应用爱因斯坦质能方程 $E=mc^2$ 可以求得地球的静止能量。

19. 请利用漫射理论估算生物组织的透射光。假设某个 10mW 的各向同性点光源被置于无限大的混浊介质中，介质的吸收系数为 0.1cm^{-1}，散射系数为 100cm^{-1}，散射的各向异性因子为 0.9，光的波长是 0.5μm。请计算距离光源 5cm 处的光能流率。

20. 请利用漫射理论估算声致发光所产生的光在生物组织中的传播。假设某个单位功率的各向同性点光源被置于无限大的混浊介质中，介质的吸收系数是 0.1cm^{-1}，散射系数为 100cm^{-1}，散射的各向异性因子是 0.9。请计算以光源为中心的半径为

5cm的整个球面上的总光能流率的积分。

21. 与只能测量漫射光幅值的连续波技术相比，频域成像技术中的相位测量能够提供什么其他信息?

推荐阅读资料

Boisdé, G. and Harmer, A. (1996). *Chemical and Biochemical Sensing With Optical Fibers and Waveguides*. Artech, Boston.

Born, M. and Wolf, E. (1999). *Principles of Optics*, 7th Expanded Ed. Cambridge Univ. Press, Cambridge, UK.

Culshaw, B. and Dakin, J. (1989). *Optical Fiber Sensors: Vol. I &II*. Artech, Boston.

Duck, F.A. (1990). *Physical Properties of Tissue*. Academic, London.

Hecht, E. (1987). *Optics*. Addison-Wesley, Reading, MA.

Katzir, A. (1993). *Lasers and Optical Fibers in Medicine*. Academic, San Diego, CA.

Lakowicz, J. (1999). *Principles of Fluorescence Spectroscopy*, 2nd Ed. Kluwer Academic/Plenum Publishers, New York, NY.

Pedrotti, F.L. and Pedrotti, L.S. (1993). *Introduction to Optics*, 2nd Ed. Prentice Hall, Upper Saddle River, NJ.

Tuchin, V. (2000). *Tissue Optics: Light Scattering Methods and Instruments for Medical Diagnosis*. Society of Photo-Optical Instrumentation Engineers Press, Bellingham, WA.

Vo-Dinh, T. (2003). *Biomedical Photonics Handbook*. CRC, Boca Raton, FL.

Welch, A.J. and van Gemert, M.J.C. (1995). *Optical-Thermal Response of Laser Irradiated Tissue*. Plenum, New York & London.

Willard, H.H., Merritt, L.L., Jr., Dean, J.A. and Settle, F.A., Jr. (1988). *Instrumental Methods of Analysis*, 7th Ed. Wadsworth, Belmont, CA.

附　　录

John Enderle 博士

A.1　MATLAB 简介

MATLAB 是一种用于科学计算和作图的高级计算机软件，其基础是矩阵运算，它的许多功能与 FORTRAN、C 等程序设计语言相同，但 MATLAB 的使用更方便。本附录介绍求解本书例题和习题所需的 MATLAB 编程知识，其中某些 MATLAB 功能已在相应章节中有过介绍。

A.1.1　矩阵的基础知识

MATLAB 最早是作为矩阵工具开发的，其多数指令都表示为矩阵的形式。矩阵（matrix）是含有 n 行 m 列元素的数组。数组用大写黑体字母表示，例如，矩阵 $\mathbf{A}$ 表示为

$$\mathbf{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \tag{A.1}$$

其中，数组的元素 a_{ij} 为常数或函数。矩阵 $\mathbf{A}$ 的阶表示为 $n\times m$，如果 $n = m$，则 $\mathbf{A}$ 为 n 阶方阵。本书所涉及的矩阵元素通常是系统参数，如电阻、电容、电感等的值。

只包含一列元素的矩阵称为列向量，它是一个 $n\times 1$ 阶数组，用小写黑体字母表示，例如：

$$\mathbf{x}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{A.2}$$

其中，各个元素 x_i 为变量、常数或者时间函数。列向量常用于表示系统的变量和输入。只包含一行元素的矩阵称为行向量，它是一个 $1\times n$ 阶数组。行向量常用于表示系统特征函数的多项式（后面有定义）。1×1 阶矩阵就是标量。

只有同阶矩阵才能进行加法和减法运算，这两种运算是将矩阵中相应的元素进行加、减运算，即：

$$
\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nm} \end{bmatrix}
$$

$$
= \begin{bmatrix} c_{11}=a_{11}+b_{11} & c_{12}=a_{12}+b_{12} & \cdots & c_{1m}=a_{1m}+b_{1m} \\ c_{21}=a_{21}+b_{21} & c_{22}=a_{22}+b_{22} & \cdots & c_{2m}=a_{2m}+b_{2m} \\ \vdots & \vdots & & \vdots \\ c_{n1}=a_{n1}+b_{n1} & c_{n2}=a_{n2}+b_{n2} & \cdots & c_{nm}=a_{nm}+b_{nm} \end{bmatrix} \tag{A. 3}
$$

加法运算时 $c_{ij}=a_{ij}+b_{ij}$；减法运算时 $c_{ij}=a_{ij}-b_{ij}$。

只有当第一个矩阵的列数等于第二个矩阵的行数时，这两个矩阵才能进行乘法运算。$n\times m$ 阶矩阵 $\boldsymbol{A}$ 乘以 $m\times n$ 阶矩阵 $\boldsymbol{B}$ 所得到的矩阵 $\boldsymbol{C}$ 为 $n\times n$ 阶矩阵，$\boldsymbol{C}$ 中的每个元素 c_{ij} 等于 $\boldsymbol{A}$ 的第 i 行元素与 $\boldsymbol{B}$ 的第 j 列元素的乘积之和，即

$$
c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \tag{A. 4}
$$

其中，$i=1, 2, \ldots, n$；$j=1, 2, \ldots, m$。一般矩阵相乘是不可交换的，即：$\boldsymbol{AB}\neq\boldsymbol{BA}$。矩阵与标量 α 的乘积等于矩阵的每个元素乘以 α。

对角线元素均为 1、非对角线元素均为 0 的方阵称为单位矩阵 $\boldsymbol{I}$（identity matrix），即

$$
\boldsymbol{I} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{A. 5}
$$

所有元素均为 0 的矩阵称为零矩阵 $\boldsymbol{0}$（null matrix）。

矩阵的加、减、乘、除运算基本遵循实数运算的规则，对于符合要求的任意矩阵 $\boldsymbol{A}$、$\boldsymbol{B}$、$\boldsymbol{C}$，有

交换律：$\boldsymbol{A}+\boldsymbol{B}=\boldsymbol{B}+\boldsymbol{A}$

结合律：$\boldsymbol{A}+(\boldsymbol{B}+\boldsymbol{C})=(\boldsymbol{A}+\boldsymbol{B})+\boldsymbol{C}$

分配律：$\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C})=\boldsymbol{AB}+\boldsymbol{AC}$

单位矩阵 $\boldsymbol{I}$ 和零矩阵 $\boldsymbol{0}$ 的性质有：

$\boldsymbol{AI}=\boldsymbol{A}$

$\boldsymbol{0A}=\boldsymbol{0}$

$\boldsymbol{A0}=\boldsymbol{0}$

$\boldsymbol{A}+\boldsymbol{0}=\boldsymbol{A}$

方阵 $\boldsymbol{A}$ 的逆记为 $\boldsymbol{A}^{-1}$，其定义是

$$
\boldsymbol{A}^{-1}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^{-1}=\boldsymbol{I} \tag{A. 6}
$$

矩阵的逆在求解系统的方程组时具有重要的作用。

矩阵 $\boldsymbol{A}$ 的转置记为 $\boldsymbol{A}'$ 或者 $\boldsymbol{A}^T$，就是将 $\boldsymbol{A}$ 的行与列进行交换，式（A.1）的转置 $\boldsymbol{A}^T$ 为

$$\boldsymbol{A}^T=\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{bmatrix}$$

方程组和矩阵

系统分析过程中经常产生含有未知变量的方程组，另外，利用初始条件求解微分方程系数时也会产生含有未知变量的方程组。

假设 i_1、i_2 和 i_3 为电路中 3 条支路的电流，并且根据基尔霍夫电压定律得到如下描述该电路的方程组：

$$\begin{aligned} 2i_1-i_2+0i_3&=6 \\ -2i_1+3i_2+i_3&=0 \\ -i_1+5i_2-4i_3&=0 \end{aligned} \tag{A.7}$$

其中，与电流相乘的常数是电路中的电阻，6 是电路的输入电压。该方程组中的每个方程都依次包含了所有 3 个变量，有的变量的系数为 0。将该方程组写成矩阵形式为

$$\begin{bmatrix} 2 & -1 & 0 \\ -2 & 3 & 1 \\ -1 & 5 & -4 \end{bmatrix}\begin{bmatrix} i_1 \\ i_2 \\ i_3 \end{bmatrix}=\begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} \tag{A.8}$$

或者为

$$\boldsymbol{Ai}=\boldsymbol{F} \tag{A.9}$$

其中

$$\boldsymbol{A}=\begin{bmatrix} 2 & -1 & 0 \\ -2 & 3 & 1 \\ -1 & 5 & -4 \end{bmatrix},\ \boldsymbol{i}=\begin{bmatrix} i_1 \\ i_2 \\ i_3 \end{bmatrix},\ \boldsymbol{F}=\begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$$

为了求解方程组（A.8）的电流向量 $\boldsymbol{i}$，将上式两边同乘以 $\boldsymbol{A}$ 矩阵的逆 $\boldsymbol{A}^{-1}$，则

$$\boldsymbol{A}^{-1}\boldsymbol{Ai}=\boldsymbol{A}^{-1}\boldsymbol{F} \tag{A.10}$$

即

$$\boldsymbol{i}=\boldsymbol{A}^{-1}\boldsymbol{F} \tag{A.11}$$

A.1.2 MATLAB 的使用

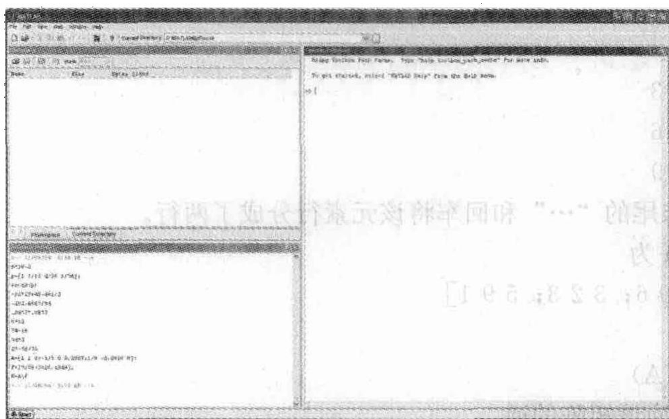
MATLAB 是一种用于数学计算并作图显示结果的软件。它配备有适用于各种工程领域的很多工具箱，包括神经网络、图像处理、数字信号处理、控制理论、线性代数、信号与系统、数值方法、符号数学以及 SIMULINK 等。SIMULINK 提供了复杂微分—积分方程的近似迭代求解方法。MATLAB 的作图功能可以满足多数用户的要求。如果需要数据统计方面的专用图形显示，那么还可以将 MATLAB 的计算结果导入到 Microsoft 的 EXCEL 软件中进行处理。

这里只简单介绍MATLAB的基本内容，读者如果需要深入学习这个软件，可以在图书馆或书店找到许多详细介绍MATLAB软件的书籍。

启动 MATLAB

单击 Microsoft Windows 的“开始”菜单，选择“程序”中的“MATLAB”项，单击其中的“MATLAB”，就可以启动MATLAB，出现如图A.1所示的界面，其中包含3个部分：指令窗、浏览器和历史指令窗。

现在我们可以在图A.1右侧的MATLAB指令窗中进行操作。在提示符“≫”后键入指令，按下回车键，就可以交互式地执行MATLAB指令。注意，MATLAB区分大小写字母，“a”与“A”是不同的，并且，MATLAB中不能像A.1节所述那样用斜体和黑体字符。



图A.1　包含3个窗的MATLAB环境（即浏览器、指令窗和历史指令窗）

MATLAB可以执行标量运算和矩阵运算。例如，键入：

≫A=10/5

MATLAB返回

A=

　2

输入矩阵时，先键入矩阵名、等号，后跟左方括号，输入数组的所有元素，再加上右方括号即可。数组元素一行行输入，先输入第一行。每行中的元素用空格或者逗号隔开，行与行之间用分号或者回车隔开。例如，键入如下矩阵 **A**：

≫A= [5 9 6; 3 2 3; 5 9 1]

MATLAB则返回

A=

```
5 9 6
3 2 3
5 9 1
```

矩阵元素也可以是函数，如：

```
≫A=[-5/2,sqrt(4),3*(6-4)]
```

MATLAB 则返回

```
A=
    -2.5000   2.0000   6.0000
```

注意，分号“;”表示元素行的结束，如果在输入行的末尾加“…”并按回车键，那么，元素就可以延续到下一行。例如：

```
≫C=[1...
2 3
4 5 6
7 8 9]
C=
    1   2   3
    4   5   6
    7   8   9
```

本例中第一行末尾的“…”和回车将该元素行分成了两行。

如果矩阵 **A** 为

```
≫A= [5 9 6; 3 2 3; 5 9 1]
```

则键入

```
≫B=inv (A)
```

MATLAB 就返回该矩阵的逆

```
B=
    -0.2941     0.5294      0.1765
     0.1412    -0.2941      0.0353
     0.2000     0          -0.2000
```

MATLAB 有指数、π 等许多特殊函数。例如，如果要用 π 进行计算，只要键入“pi”，即：

```
≫pi
ans=
     3.1416
```

对于每条输入指令，MATLAB 都会返回一个数值结果。指数函数的调用方法是“exp (-0.5)”，键入以后，就可以得到其输出为 0.6065。正弦和余弦等高级计算器中具有的基本标准内置函数在 MATLAB 中都有。打开 MATLAB 的帮助窗，再选择“Functions-Alphabetical List”，就可以了解所有 MATLAB 函数。

A.1.3 算术表达式

MATLAB 的算术运算符号如下：

^　*　/　\　+　-

它们依次为幂运算、乘、右除、左除、加和减。注意，1/4 和 4\1 分别表示 1 除以 4 和 4 除 1，都等于 0.25。如下面的例 A.1 所示，我们常用左除“\”运算符求解方程组。

这些运算符在 MATLAB 中的计算顺序就是这里的排列次序，即“^”最先，其次是“*”、“/”、“\”，然后是“+”、“-”。例如：

```
>>5+6^2*3/7*2
```

输出结果是

```
ans=
    35.8571
```

该式首先计算 6^2=36，然后 36 乘 3，再除以 7，乘以 2，最后加 5。与数学运算相似，计算由左至右，使用圆括号可以改变计算次序。例如：

```
>>3*(6-4)
```

输出结果是

```
ans=
    6
```

该式首先计算 6-4，再将结果乘以 3。

MATLAB 的三角函数计算中的角度用弧度表示，例如，求 sin (45°)，要键入

```
>>sin(pi/4)
```

输出结果是

```
ans=
    0.7071
```

例 A.1　请求解如下方程组：

$$
\begin{gathered}
4i_1-i_2-i_3-i_4-i_5=240\\
-i_1+8i_2-i_3+0\times i_4+0\times i_5=0\\
-i_1-i_2+5i_3-i_4+0\times i_5=0\\
-i_1+0\times i_2-i_3+5i_4-i_5=0\\
-i_1+0\times i_2+0\times i_3-i_4+8i_5=0
\end{gathered}
$$

解：将方程组写成矩阵形式

$$\boldsymbol{Ai}=\boldsymbol{F}$$

其中

$$
\boldsymbol{A}=\begin{bmatrix}4&-1&-1&-1&-1\\-1&8&-1&0&0\\-1&-1&5&-1&0\\-1&0&-1&5&-1\\-1&0&0&-1&8\end{bmatrix};\ \boldsymbol{i}=\begin{bmatrix}i_1\\i_2\\i_3\\i_4\\i_5\end{bmatrix};\ \boldsymbol{F}=\begin{bmatrix}240\\0\\0\\0\\0\end{bmatrix}
$$

使用左除“\”运算，可以方便地求得列向量 $\boldsymbol{i}$，即

$$\boldsymbol{A}\backslash\boldsymbol{Ai}=\boldsymbol{i}=\boldsymbol{A}\backslash\boldsymbol{F}$$

MATLAB 求解 i 向量的指令如下：

```
≫A=[4 -1 -1 -1 -1;-1 8 -1 0 0;-1 -1 5 -1 0;-1 0 -1 5-1;-1 0 0-1 8];
≫F=[240;0;0;0;0;];
≫i=A\F
i=
   77.5000
   12.5000
   22.5000
   22.5000
   12.5000
```

■

A.1.4 向量

MATLAB 以行向量的形式存储向量，第一个元素为 i 分量，第二个元素为 j 分量，第三个元素为 k 分量，都写在方括号里。例如，向量 $\boldsymbol{F}=3\boldsymbol{i}-4\boldsymbol{j}+2\boldsymbol{k}$ 在 MATLAB 中的键入方式为

```
≫F= [3 -4 2]
```

结果是

```
F=
   3 -4 2
```

MATLAB 指令 “norm” 用于计算向量的幅值，例如，向量 $\boldsymbol{F}$ 的幅值为

```
≫Fmagnitude=norm (F)
Fmagnitude=
              5.3852
```

计算两个向量点积的 MATLAB 指令为 “dot (A, B)”。例如，假设 $\boldsymbol{A}=5\boldsymbol{i}-8\boldsymbol{j}-2\boldsymbol{k}$，$\boldsymbol{B}=6\boldsymbol{i}+4\boldsymbol{j}-3\boldsymbol{k}$，在 MATLAB 中计算两者的点积：

```
≫A= [5 -8 -2];
≫B= [6 4 -3];
≫ABdot=dot (A, B)
ABdot=
          4
```

计算两个向量的叉积的 MATLAB 指令为 “cross (A, B)”，它也是一个内置函数。例如，$\boldsymbol{A}\times\boldsymbol{B}$ 为

```
≫A= [5 -8 -2];
≫B= [6 4 -3];
≫ABcross=cross (A, B)
ABcross=
          32   3   68
```

A.1.5　复数

MATLAB中所有的数都作为复数存储，用 i 或 j 表示复数的虚部，也就是在MATLAB中 $z=3+i4$ 和 $z=3+j4$ 是等价的，即

```
≫z=3+4j
```

结果是

```
z=
   3.0000+4.0000i
```

而

```
≫z=3+4i
```

结果也是

```
z=
   3.0000+4.0000i
```

如前所述，本书中用符号 i 表示电流，用符号 j 表示复数。但是，MATLAB的默认复数符号是 i。MATLAB中的复数也可以用极坐标形式来表示，即对于任意实数 x 和 y，有

$$z=x+jy=|z|e^{j\theta}$$

其中

$$|z|=\sqrt{x^2+y^2}$$

$$\theta=\arctan\frac{y}{x}$$

MATLAB有很多处理复数的内置函数，表A.1列出了其中几个基本函数。

表A.1　复数的MATLAB函数指令（$z=x+jy$）

数　值	指　令
Re（z）	real（z）
Im（z）	imag（z）
$\|z\|=\sqrt{x^2+y^2}$	abs（z）
$\theta=\arctan\frac{y}{x}$	angle（z）

例A.2　假设 $z_1=3+j2$，$z_2=2-j$，请用MATLAB求解 $z_3=\frac{z_1}{z_2}$。

解： MATLAB求解过程如下：

```
≫z1=3+j*2
z1=
   3.0000+2.0000i
≫z2=2-j
z2=
```

```
    2.0000－1.0000i
≫z3＝z1/z2
z3＝
    0.8000＋1.4000i
```
■

例 A.3 请计算：$z=(60+j30)+100e^{-j0.4887}$。

解：用如下 MATLAB 指令，混合使用复数的直角坐标和极坐标两种形式，计算 z 值：

```
≫z＝60＋j＊30＋100＊exp（－j＊0.4887）
z＝
  1.4829e＋002－1.6948e＋001i
```
■

例 A.4 试计算如下复数的幅值和角度：

$$z=(5-j4)^3$$

解：
```
≫z＝(5－j＊4)^3
z＝
    －1.1500e＋002－2.3600e＋002i
≫r＝abs(z)
r＝
    262.5281
≫theta＝angle(z)
theta＝
        －2.0242
```
■

A.1.6 多项式及其根

MATLAB 用行向量处理多项式，并且可以用内置函数计算多项式的根。行向量中多项式系数按照阶次下降的顺序排列，最高阶次的系数在最前面。假设多项式为

$$x^4-12x^3+0x^2+25x+116=0 \tag{A.12}$$

则在 MATLAB 中表示为

```
≫p＝[1 －12 0 25 116]
```

结果是

```
p＝
  1 －12 0 25 116
```

每项系数都必须输入，系数为 0 的项也要输入。然后，用内置函数“roots”就可以求得多项式的根。方程（A.12）的求根指令为：

```
≫r＝roots(p)
```

MATLAB 返回如下两个实数根和一对共轭复数根：

```
r＝
  11.7473
```

2.7028

−1.2251+1.4672i

−1.2251−1.4672i

MATLAB 规定多项式是行向量，而根是列向量。

反过来，如果已知多项式的根，可以用 MATLAB 指令"poly"求得多项式。在如上求得的多项式根的基础上，键入"poly"指令就可以恢复原来的多项式，即

```
>>pp=poly(r)
```

返回结果

```
pp=
   1.0000  -12.0000  -0.0000  25.0000  116.0000
```

利用 MATLAB 函数"conv"可以进行多项式的乘法运算，也就是求两个多项式的卷积。例如，要将如下两个多项式相乘

$$A(x)=x^3+2x^2+3x+4 \tag{A.13}$$

$$B(x)=x^3+4x^2+9x+16 \tag{A.14}$$

就在 MATLAB 中键入

```
>>A= [1 2 3 4];
>>B= [1 4 9 16];
>>AB=conv (A, B)
```

返回结果

```
AB=
   1   6   20   50   75   84   64
```

注意，矩阵 **A** 和 **B** 后面分号的作用是取消指令的返回结果显示。这两个多项式相乘的结果为

$$AB(x)=x^6+6x^5+20x^4+50x^3+75x^2+84x+64 \tag{A.15}$$

具有相同阶数的多项式才能相加，例如，要将如下两个多项式相加

$$S(x)=x^5+4x^4+5x^3+6x^2+7x+2 \tag{A.16}$$

$$T(x)=2x^4+9x^3+8x^2-4x+5 \tag{A.17}$$

由于（A.17）这个多项式为 4 阶，因此，要添加一个系数为 0 的 x^5 项。这样，MATLAB 中输入的多项式为

```
>>S= [1 4 5 6 7 2];
>>T= [0 2 9 8 -4 5];
>>U=S+T
```

返回结果

```
U=
  1   6   14   14   3   7
```

也就是

$$x^5+6x^4+14x^3+14x^2+3x+7 \tag{A.18}$$

利用MATLAB函数“deconv（A，B)”可以进行多项式的除运算，它表示B除A，也就是A除以B。通常多项式的除都有余数，因此该函数的调用形式是［Q，V］＝deconv（A，B)，其中，Q为多项式的商，V为余数。例如，以上多项式（A.16）与（A.17）的比，也就是 $x^5+4x^4+5x^3+6x^2+7x+2$ 除以 $2x^4+9x^3+8x^2-4x+5$ 为

```
≫S= [1 4 5 6 7 2];
≫T= [2 9 8 -4 5];
≫ [Q, R] =deconv (S, T)
Q=
   0.5000-0.2500
R=
   0  0  3.2500  10.0000  3.5000  3.2500
```

将该结果写成符号形式为

$$\frac{1}{2}x-\frac{1}{4}+\frac{3.25x^3+10x^2+3.5x+3.25}{2x^4+9x^3+8x^2-4x+5}$$

下面举一个没有余数的简单例子，即 x^2-2x+1 除以 $x-1$，则

```
≫S= [1, -2, 1 ];
≫T= [1, -1];
≫ [Q, V] =deconv (S, T)
Q=
   1 -1
V=
   0  0  0
```

此结果是 $x-1$。

例A.5 设有如下方程：

$$\frac{(1600-3x)^3}{27x^3}=\frac{x}{500-x}$$

试找出0到500的 x 的值。

解：由于常规MATLAB计算表达式时使用的是“format short”，为了使数值具有足够多的有效位数，这里我们使用“format long”指令。首先计算 $(1600-3x)^3$：

```
≫format long
≫y=[-3 1600];
≫r=conv(y, y);
≫w=conv(y, r)
```

返回的结果为

```
w=
   1.0e+009 *
```

−0.00000002700000 0.00004320000000 −0.02304000000000 4.09600000000000

其中“w”为表达式（1600−3x）^{3}。注意，“1.0e+009”是 w 表达式所有系数的倍数因子。

为了将题目中方程右边的分母消去，下面将该方程两边同乘以 500−x，即：

≫z=［−1 500］;

还是用“conv”完成乘法 w * z

≫s=conv(z，w)

结果为

s=

1.0e+012 *

Columns 1 through 4

0.00000000002700 −0.00000005670000 0.00004464000000 −0.01561600000000

Column 5

2.04800000000000

方程左边分母 $27x^3$ 乘以右边分子 x 得到（$27x^4$），在 MATLAB 中是

≫q=［27 0 0 0 0］

输出为

q=

27 0 0 0 0

然后，将 s 表达式与 q 表达式相减，即

≫t=s−q

得到

t=

1.0e+012 *

Columns 1 through 4

0 −0.00000005670000 0.00004464000000 −0.01561600000000

Column 5

2.04800000000000

以上计算得到的多项式 t 就是

$$-5.67\times 10^4 x^3 + 4.464\times 10^7 x^2 + 1.5616\times 10^{10} x + 2.048\times 10^{12} = 0$$

最后，求该多项式 t 的根，即

≫v=roots(t)

得到

v=

1.0e+002 *

2.62504409150663+2.62297313283571i

2.62504409150663−2.62297313283571i

2.62292769000261

其中两个复数根不在题目给定的数值范围内，符合题目要求的解是

$$x=262.3$$

■

表 A.2 归纳了一些用于多项式计算的 MATLAB 指令。

表 A.2 MATLAB 的多项式指令

运算	符号表达式	指令
多项式	$a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0$	$P=[a_na_{n-1}\cdots a_1a_0]$
多项式求根	$(x-r_1)(x-r_2)\cdots(x-r_n)$	r=roots(P)
根据根写出多项式	$a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0$	poly(r)
两个多项式的乘积 $S(x)=A(x)\times B(x)$	$S(x)=(a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0)\times(b_mx^m+b_{m-1}x^{m-1}+\cdots+b_1x+b_0)$	$S=$conv(A,B)
两个多项式的加和 $S(x)=A(x)+B(x)$；注意，多项式的阶次必须相同。	$S(x)=(a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0)+(0x^n+\cdots+b_mx^m+b_{m-1}x^{m-1}+\cdots+b_1x+b_0)$	$S=A+B$
两个多项式的除 $S(x)=\frac{A(x)}{B(x)}=Q(x)+V(x)$；	$S(x)=\frac{a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0}{b_mx^m+b_{m-1}x^{m-1}+\cdots+b_1x+b_0}=q_px^p+\cdots q_1x+q_0+\frac{v_lx^l+\cdots+v_1x+v_0}{b_mx^m+b_{m-1}x^{m-1}+\cdots+b_1x+b_0}$	[Q,V]=deconv(A,B)

A.1.7 MATLAB 作图

MATLAB 具有很强的二维和三维向量（即矩阵）的作图功能，可以用于数据的可视化图形显示，并且根据用户的需要可以设计不同的图形显示方式、线宽、颜色、注释和标题等。下面介绍几种作图指令。

fplot 指令

MATLAB 的基本作图指令是 “fplot（‘f’，[T_1，T_2]）”，它显示 $T_1\sim T_2$ 区间上的 f 函数曲线。注意，表达式 “f” 放在单引号内，这样，MATLAB 就会将其作为字符串表达式，而不是计算 “t” 时刻的函数值。例如，要作出 $f=2e^{-t}$ 在 0 到 5 区间上的曲线图，MATLAB 指令为

```
>>fplot('2 * exp(-t)',[0,5])
```

该指令将函数 f 在 0～5 时间区间上的曲线显示在如图 A.2 所示的独立窗口中。如果要编辑图形，选中打印图标旁边的箭头图标，然后双击图形，就可以打开另一个指令窗，在此可以修改线条的颜色、线形、字体、坐标轴标题、坐标轴等图形的属性。

MATLAB 图形可以通过剪贴板方便地复制和粘贴到 Microsoft Word 等其他应用软件中。利用 MATLAB 图形窗的编辑菜单还可以控制图形的复制方式，例如，“Metafile” 格式复制图形时无背景颜色，而 “Bitmap” 格式则完全复制图形，包括背景颜色。

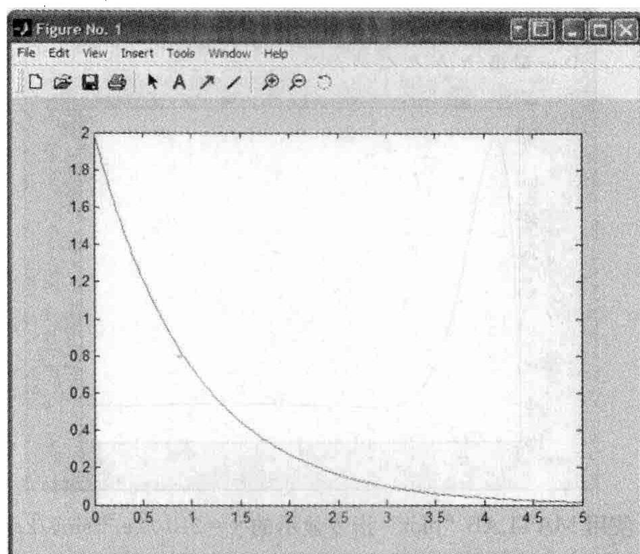


图 A.2　使用 MATLAB“fplot”指令显示的 $f=2e^{-t}$ 函数曲线

plot 指令

另一条作图指令是“plot(t，y)”，它作出 y 对 t 的变化曲线，数据点之间用直线连接。用“linspace”指令可以定义自变量 t，该指令的写法是“t=linspace(a，b，n)”。它产生含有 n 个元素的行向量，元素值线性分布于 a～b 的区间上。

例如，如下指令可以作出函数 $y=210\sqrt{2}e^{-2t}\sin(\sqrt{2}t)$ 的曲线：

```
≫t=linspace(0,8,1000);
≫y=297 * sin(1.414 * t). * exp(-2 * t);
≫plot(t,y)
```

其中，第一条指令创建含有 1000 个元素的行向量 t，元素的值分布于 0～8 区间上。第二条指令通过“. *”运算将 sin(1.414 * t) 与 exp(-2 * t) 相乘，创建含有 1000 个元素的行向量 y。“. *”运算是点积运算，它执行矩阵中对应元素之间的相乘。在本例的这个运算式中，如果误用“*”运算，而不是“. *”运算，MATLAB 就会出错。因为此处 t 为 1×1000 的行向量，使得 sin（1.414 * t）和 exp（-2 * t）都是 1×1000 的行向量，这样，两个矩阵的阶不符合矩阵乘法的要求。图 A.3 显示了以上“plot”指令所作的 0～8 时间区间上 1000 个点连接而成的 y 函数曲线。

用“plot”指令还可以在同一张图上画出多条函数曲线。例如，如下是将 $f=100e^{-t}$ 和 $y=210\sqrt{2}e^{-2t}\sin(\sqrt{2}t)$ 两个函数显示在同一张图上的指令：

```
≫t=linspace(0,8,1000);
≫y=297 * sin(1.414 * t). * exp(-2 * t);
≫f=100 * exp(-t);
≫plot(t,y,t,f)
```

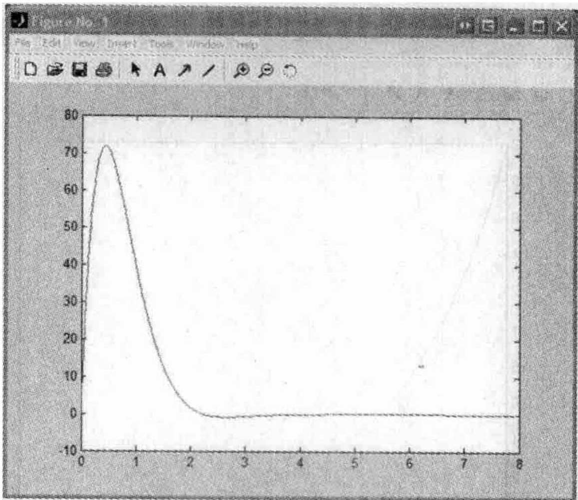


图 A.3 使用 MATLAB “plot” 指令显示的 $y=210\sqrt{2}e^{-2t}\sin(\sqrt{2}t)$ 函数曲线

生成的图形如图 A.4 所示。如上所述，根据用户的需要可以改变图形的各项属性，包括线条的颜色、线形、字体、坐标轴及其标题等等。

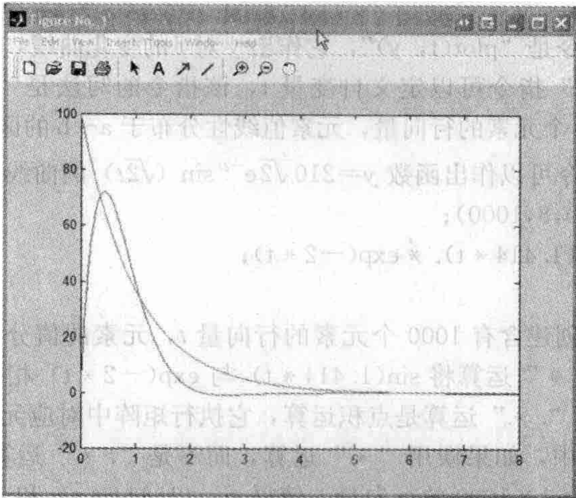


图 A.4 使用 MATLAB “plot” 指令显示的 $f=100e^{-t}$ 和 $y=210\sqrt{2}e^{-2t}\sin(\sqrt{2}t)$ 函数曲线

MATLAB 的作图指令还有很多，包括三维图、直方图、饼形图等等。读者可以在 MATLAB 的帮助信息中查阅所有作图函数。

Microsoft Excel 作图

虽然 MATLAB 提供了很强的作图功能，但是，如果将数据导出到 Microsoft Excel 等其他应用软件，可以绘制出更专业化的图形。这里，首先要把数据从 MATLAB 中导出，然后导入到 Excel 中。MATLAB 的所有数据都存放在工作空间中，各种工具箱（包括 SIMULINK）都可以通过变量名访问数据。用 MATLAB 的 “who” 或

"whos" 指令可以列出工作空间中存在的所有变量。"who" 指令只列出变量名，"whos" 指令列出变量名、大小、所用字节数以及变量的类型（即逻辑字符、整数数组、浮点数数组等）。

在数据导出之前，先要把数据组合到同一个数组矩阵中，自变量在前，后跟各个数据点上对应于自变量计算得到的因变量。例如，如果要导出图 A.3 所示函数的数据，t 和 y 的值为

```
>>t=linspace(0,8,1000);
>>y=297 * sin(1.414 * t). * exp(-2 * t);
```

记住，此处 t 和 y 存放成 1×1000 的行向量，因此，必须先将这两个行向量转置，才能组合成新的数据矩阵 ***data***，即：

```
>>data=[t',y'];
```

该 ***data*** 矩阵为 1000×2 阶。如果如图 A.4 所示还有另外向量要一起导出，那么，可以用 "data= [t', y', f'];" 指令组合数据。

如果要将数据导出到 ASCII 码文件 "output. txt" 中，并且用制表符 tab 分隔变量值，那么可以用如下指令：

```
>>save output. txt data-ascii-tabs
```

该文件存放在默认的 MATLAB 子目录 "work" 中。现在，就可以在 Excel 中打开 "output. txt" 文件了。打开之后，数据存放在 Excel 工作空间的前两列中，按照作图向导的指示，就可以一步步创建出如图 A.5 所示的散点图。

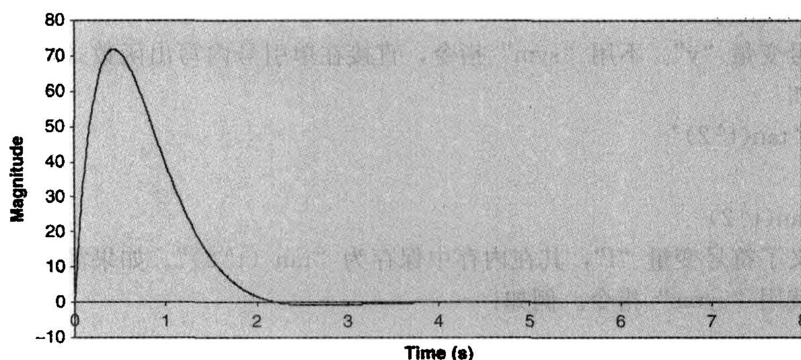


图 A.5　利用 Microsoft Excel 作出的 $y=210\sqrt{2}e^{-2t}\sin(\sqrt{2}t)$ 函数曲线

A.1.8　工作空间的加载和保存

许多问题可以用 MATLAB 的交互运行方式求解，也就是在指令行中输入信息，并按回车键，MATLAB 就会给出表达式的计算结果。如果问题比较复杂，程序较长，就要用 m 文件保存程序，以便以后可以随时调用。用户逐渐精通 MATLAB 之后，就会经常使用 m 文件的 MATLAB 指令、数据和输入输出信息。MATLAB 数据文件的默认扩展名是 ".mat"。保存工作空间可以生成 m 文件，或者在 Windows 记事本中输入指令并保存为 m 文件，再在 MATLAB 中用 "load" 指令加载这些指令，

加载时，m 文件中的所有表达式都会被计算求值。

打开 m 文件的指令是

≫load"filename. mat"

"save" 指令用于将工作空间保存到 m 文件中。如果没有给定文件名，MATLAB 就用默认文件名 "matlab. mat"。

A.2 利用 MATLAB 求解微分方程

MATLAB 提供符号数学工具箱，可以用符号法求解各种微积分问题和代数问题。该工具箱是基于名为 Maple 软件的一组函数运算指令集，Maple 由加拿大 Ontario 的 University of Waterloo 大学开发。与前面所介绍的 MATLAB 指令不同，这个工具箱指令的运行结果不是数值，而是符号函数。下面是几个符号函数的例子：

$$\tan(t^2)，\frac{d}{dt}(4y^2+2y+4)，\int\sin(y)e^{-2y}dy \text{ 和 } 4\ddot{y}+2\dot{y}+6=e^{-2t}。$$

符号数学工具箱里的 MATLAB 指令可用于求取函数的微分、积分或者微分方程的精确符号解。例如，求 $\tan(t^2)$ 的导数，就会得到 MATLAB 返回的解 $2t(1+\tan(t^2)^2)$。下面将介绍用于本附录问题求解的一些指令。

执行符号运算之前先必须定义所有符号变量，符号变量用 MATLAB 的 "sym" 指令或者 "syms" 指令定义，变量名写在单引号内。例如，以下指令

```
>>y=sym('y')
```

建立了符号变量 "y"。不用 "sym" 指令，直接在单引号内写出函数式也可以定义符号变量，如

```
>>f='tan(t^2)'
  f=
    tan(t^2)
```

该指令定义了符号变量 "f"，其在内存中保存为 "tan (t^2)"。如果需要定义多个符号变量，就用 "syms" 指令。例如：

```
>>syms x y w z
```

或者写成

```
>>syms('w','x','y','z')
```

这两条指令都定义了 4 个符号变量，它们与用 "sym" 指令分别定义单个变量的结果相同。任何函数如果使用已定义的符号变量，那么，该函数就是符号函数。例如：

```
>>f=tan(y^2)
    f =
      tan (y^2)
```

注意，由于 "y" 已经被定义为符号变量，该 MATLAB 指令中不需要用单引号写成 f='tan(t^2)'。这些已定义的符号变量 "w"、"x"、"y" 和 "z" 就可以用于求取符

号计算结果，例如，可以用 MATLAB 指令“det”建立符号行列式，即

```
>>m=[w,x;y,z]
      m=
        [w,x]
        [y,z]
>>det(m )
        ans=
           w*z-x*y
```

MATLAB 指令“diff”和“int”分别为符号求导和符号积分。下面是这两个符号函数的应用例子：

```
>>x=diff('cos(y)')
      x=
        -sin(y)
>>x=int('sin(y)')
      x=
        -cos(y)
```

如果 MATLAB 符号计算的结果难以辨认，可以用“simplify”指令进行代数化简，生成比较容易识别的形式。例如：

```
>>simplify((x^2+7*x+12)/(x+4))
  =
    x+3
```

下面大家就会看到，这条指令在积分和微分运算中很有用。

在 MATLAB“int”指令中加入积分限，还可以计算定积分。例如，以下指令计算 $\int_0^{\pi}\cos(y)dy$ 的积分：

```
>>x=int('sin(y)','0','pi')
      x=
         2
```

积分限并不一定要是数值，也可以是符号变量，例如

```
>>x=int('sin(y)','a','b')
      x=
        -cos(b)+cos(a)
```

用“solve”指令求解符号代数表达式很方便，例如，以下指令求多项式 $x^4+14x^3+71x^2+154x+120=0$ 的根：

```
>>solve('x^4+14*x^3+71*x^2+154*x+120=0')
ans=
    [-5]
```

```
[-4]
[-3]
[-2]
```

前面讲过，用“roots”指令可以求多项式的根，即

```
>>p= [1 14 71 154 120];
>>r=roots(p)
  r =
    -5.0000
    -4.0000
    -3.0000
    -2.0000
```

这两条指令的区别在于“solve”指令求根的符号解时不需要给定多项式系数的值。例如，以下指令求众所周知的二次方程的根：

```
>>syms a b c y x
>>solve('a * x^2+b * x+c=0')
ans=
    [1/2/a * (-b+(b^2-4 * a * c)^(1/2))]
    [1/2/a * (-b-(b^2-4 * a * c)^(1/2))]
```

为了求解动态系统的输出响应，通常需要求解包含数个未知变量（如自由响应中的系数）的系统方程组。例如，求解例 8.17 时，就需要求 $K_1+K_2=3$ 和 $-5K_1-3K_2=299$ 组成的方程组的解，当时用了矩阵方法。这里，也可以用以下“solve”指令求解这个方程组：

```
>>syms k1 k2
>> [k1, k2] =solve('k1+k2=3','-5 * k1-3 * k2=299')
      k1=
          -154
      k2=
          157
```

“solve”指令还有许多其他调用方式，读者可以查阅 MATLAB 的帮助信息或者参考其他书籍深入了解该指令的使用方法。

求解常微分方程的 MATLAB 指令是“dsolve”，该指令中用大写字母“D”表示一阶导数，用“D2”表示二阶导数，用“D3”表示三阶导数，依此类推。指令的参数包括微分方程和初始条件，用逗号隔开，并分别写在单引号内。例如，如果用“dsolve”指令求解如下微分方程在 $t\geqslant 0$ 时的解：

$$\ddot{y}+4\dot{y}+3y=0$$

初始条件为 $y(0)=1$ 和 $\dot{y}(0)=0$，则指令如下

```
>>dsolve('D2y+4 * Dy+3 * y=0','dy(0)=0','y(0)=1')
```

ans =

　　$-1/2 * exp(-3 * t)+3/2 * exp(-t)$

MATLAB 的作图函数“ezplot”可以用于显示符号函数曲线，例如，运行指令

```
>>ezplot(y,[0,5])
```

就可以得到如图 A. 6 所示的本题的解曲线。“ezplot”指令中方括号内的参数指定了作图的起点和终点。

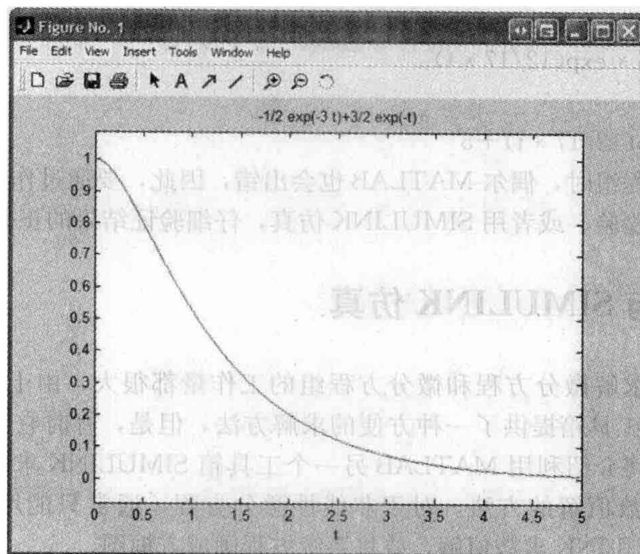


图 A. 6　MATLAB 作图函数“ezplot”的输出图形

如果“dsolve”指令中没有给定初始条件，那么，解就用未知参数表示。例如，以下指令不包含初始条件

```
>>dsolve('D2y+4 * Dy+3 * y=0')
```

返回的解就是

$$C1 * exp(-t)+C2 * exp(-3 * t)$$

其中 C1 和 C2 的值要用初始条件来确定。

例如，用 MATLAB 求解微分方程

$$\ddot{y}+16000\dot{y}+10^8 y=5\times10^8 t$$

初始条件为 $\dot{y}(0)=300$ 和 $y(0)=-45\times10^{-3}$，则

```
>>dsolve('D2y+16000 * Dy+10^8 * y=5 * 10^8 * t',
        'Dy(0)=300','y(0)=-45 * 10^(-3)')
ans=
    -1/1250+5 * t-293/30000 * exp(-8000 * t) *
  sin(6000 * t)-221/5000 * exp(-8000 * t) * cos(6000 * t)
```

利用“dsolve”指令也可以求解微分方程组，只要将方程组中的每个方程和初始条件都分别列在参数项中，用单引号括起来，并用逗号隔开就可以了。例如，如果

需要求解以下方程组：

$$5\dot{y}_2+10\dot{y}_1+60y_1=300u(t)$$

$$5\dot{y}_2+40y_2+5\dot{y}_1=0$$

$t=0$ 时的初始条件为 0。则 MATLAB 指令为

```
>>[x,y]=dsolve('5 * Dy+10 * Dx+60 * y=300',
    '5 * Dy+40 * y+5 * Dx=0','x(0)=0','y(0)=0')
x=
    -45+45 * exp(12/17 * t)
y=
    -5 * exp(12/17 * t)+5
```

求解微分方程组时，偶尔 MATLAB 也会出错，因此，要通过作图显示、初始条件和终止条件的检验、或者用 SIMULINK 仿真，仔细验证结果的正确性。

A.3 框图与 SIMULINK 仿真

如前所述，求解微分方程和微分方程组的工作量都很大，由上节可知，MATLAB 的符号数学工具箱提供了一种方便的求解方法，但是，有时它并不能给出正确的解。本节我们将介绍利用 MATLAB 另一个工具箱 SIMULINK 求线性微分方程和非线性微分方程数值解的方法。对于非线性微分方程，通常只能用 SIMULINK 求解。为了用 SIMULINK 求数值解，要将系统方程画成方框图。

A.3.1 框图

框图是系统微分方程的图形表示，它用常数、增益、加法器和积分器等基本数学计算来表示方程。处理非线性元件时，可以建立包含二次方根、指数、对数等数学函数的特殊模块。如图 A.7 显示了框图的基本元件。

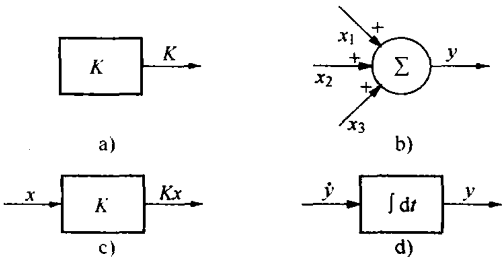


图 A.7 框图的基本元件

a) 常数模块 b) 加法器 c) 增益模块 d) 积分器

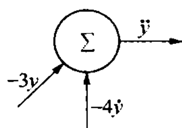
创建框图时，首先，将最高阶导数项作为输出变量放在等式左边，把其余项都放在等式右边。然后，将等式右边各项都连接到代表等式的一个加法器上，加法器的输出就是等式的左边。在加法器的输出上添加足够的积分器，把最高阶导数降为输出变量（如$\frac{d^3y}{dt^3}$需要 3 个积分器等）。最后，通过增益模块将输出变量及其各阶导数反过来连接到加法器上。如果存在输入信号，就将它们也添加到加法器上。输入可以是常数模块表示的单位阶跃函数，或者是 t、e^{-t}等函数模块。

例 A.6　请画出方程 $\ddot{y}+4\dot{y}+3y=0$ 的框图。

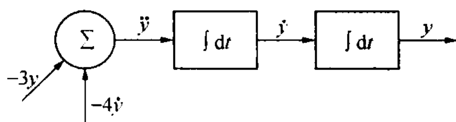
解：首先，将微分方程重排为

$$\ddot{y}=-4\dot{y}-3y$$

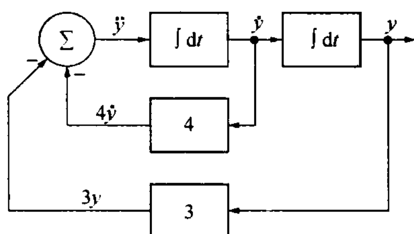
再作出如下框图的加法器：



然后，如下所示，在加法器的输出上添加两个积分器，得到 $\dot{y}$ 和 y：



最后，如下所示，通过增益模块将 $\dot{y}$ 和 y 连接到加法器上，框图就完成了。



用这种方法也可以建立微分方程组的框图，就是分别画出每个微分方程的框图，并把输出变量及其各阶导数连接到相应的加法器上。■

例 A.7　请画出如下微分方程组的框图：

$$3\ddot{y}_1+10\dot{y}_1+60y_1+5\dot{y}_2=f(t)$$

$$5y_1+4\ddot{y}_2+5\dot{y}_2+40y_2=0$$

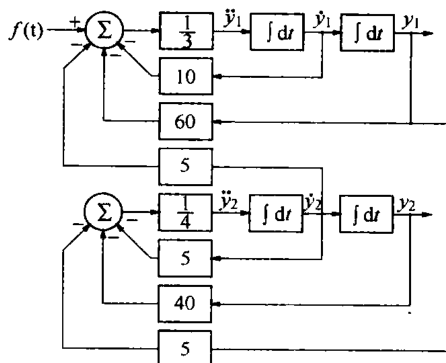
其中，$f(t)$ 为不确定的输入函数。

解：首先，求各方程的最高阶导数，即

$$\ddot{y}_1=\frac{1}{3}(-10\dot{y}_1-60y_1-5\dot{y}_2+f(t))$$

$$\ddot{y}_2=\frac{1}{4}(-5y_1-5\dot{y}_2-40y_2)$$

建立框图时，先画出两个加法器，它们的输出分别是 $\ddot{y}_1$ 和 $\ddot{y}_2$，再在每个输出变量之后添加两个积分器；然后，通过相应的增益元件将两个方程右边的各项加到加法器的输入上，就完成了如下所示的框图。



■

A.4 SIMULINK 仿真工具箱

SIMULINK 是 MATLAB 的一个工具箱，可用于仿真线性和非线性、连续和离散的动态系统。下面我们根据上一节所介绍的系统方程的框图，在 SIMULINK 中通过拖拉和连接模块建立 SIMULINK 的仿真模型，并运行仿真程序，作图输出结果，或将输出结果传送给 MATLAB。

启动 SIMULINK 时，先启动 MATLAB，在指令窗中键入小写“simulink”，就可以打开如图 A.8 所示的另一个包含各种 SIMULINK 仿真图标的窗口。然后，选择菜单栏中的“File”菜单下的“New”和“Model”项。这样，又打开一个新的窗口，它是如图 A.9 所示的用于建立模型的 SIMULINK 工作空间。图 A.8 中的各种图标代表用于仿真建模的基本模块子系统，每个子系统含有一组模块，构成模块库，模块库是功能类似的模块的集合。

双击图标可以打开模块库，例如，双击数学运算“Math Operations”模块库打开的窗口如图 A.10 所示，其中包含了求绝对值、增益、数学函数等常用计算。连续模块库“Continuous”包含有求导数和求积分等模块。信号源模块库“Sources”包含有常数、时钟、阶跃函数等模块，这些模块都有一个输出，但没有输入。时钟“clock”模块输出当前仿真时间变量，它在选择模块以及向 MATLAB 输出变量时很有用。接收器模块库“Sinks”中的模块都只有一个输入，没有输出。例如，用于显示变量随时间变化的示波器模块就是其中的一个模块。这个模块库中还有一个重要的模块是“To Workspace”模块，它可以将变量传送到 MATLAB 的工作空间中，以便进一步分析处理和作图显示。

直接将 SIMULINK 库浏览器“SIMULINK Library Brower”窗口中的模块拖拉到图 A.9 所示的工作空间中就可以复制各种模块。每个模块放到工作空间之后，就有了特定的名称，显示在模块下方。如果有多个同样的模块，其名称就是模块名后加一个整数。模块内部还有一个模块标识符。将鼠标指向模块名，单击鼠标左键，就可以编辑模块名。双击工作空间中的模块，可以打开一个新的窗口，用于设置模

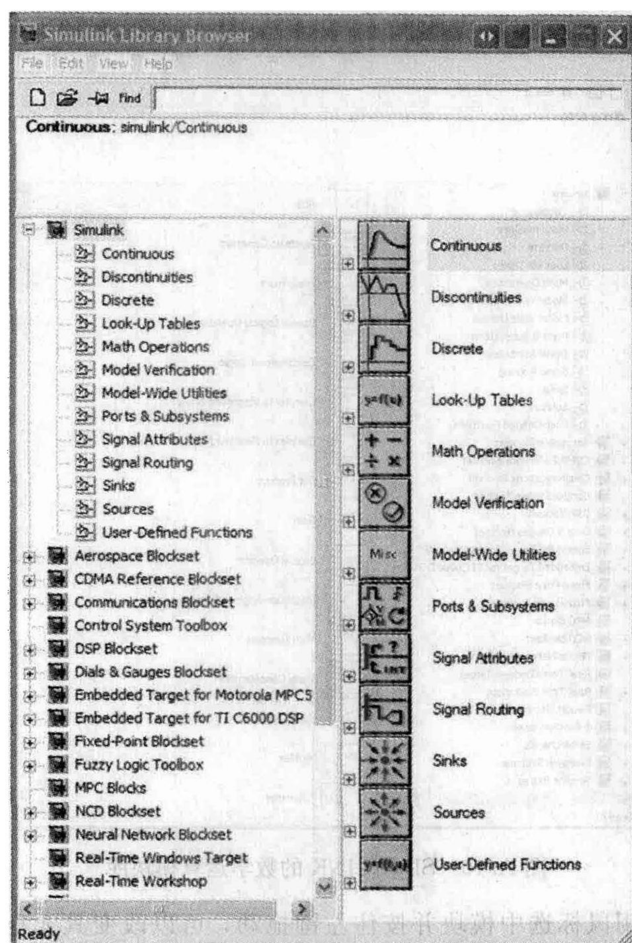


图 A.8　SIMULINK 模块库（右边）和 MATLAB 的工具箱（左边）

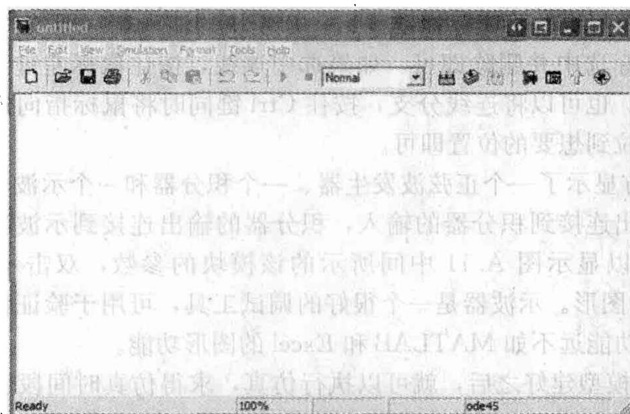


图 A.9　用于建立仿真模型的 SIMULINK 工作空间

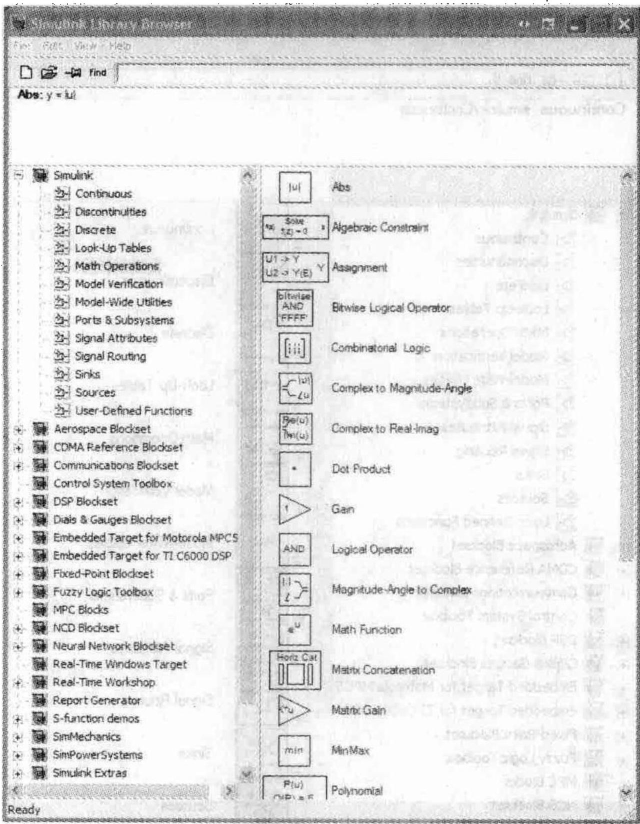


图 A.10 SIMULINK 的数学运算模块库

块的各种参数。用鼠标选中模块并按住左键拖动，可以改变其大小。要去除模块，只要选中并删除即可。为了方便，可以用复制和粘贴操作复制模块。

模块之间用直线段连接。连接两个模块时，将鼠标指向一个模块的输出端（即“>”处)，按下鼠标键并将光标拖动到另一个模块的输入端（即“<”处)。如果连线不对，就用鼠标选中并删除即可。如果选中连线，按住鼠标左键，可以将连线拖拉成所需的形状。也可以将连线分支，按住 Ctrl 键同时将鼠标指向连线并按住鼠标左键，将光标拖拉到想要的位置即可。

图 A.11 上方显示了一个正弦波发生器、一个积分器和一个示波器组成的仿真模型，正弦波的输出连接到积分器的输入，积分器的输出连接到示波器的输入。双击正弦波发生器可以显示图 A.11 中间所示的该模块的参数，双击示波器则显示图 A.11 下方所示的图形。示波器是一个很好的调试工具，可用于验证仿真输出是否正确，但是其作图功能远不如 MATLAB 和 Excel 的图形功能。

SIMULINK 模型建好之后，就可以执行仿真，求得仿真时间段上的输出和输入结果。在运行仿真之前，先要单击菜单上的仿真“Simulation”项，选择下拉菜单中

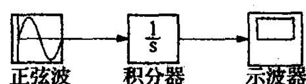
的仿真参数“Simulation Parameters”项，设置仿真参数。图 A.12 所示是仿真参数窗口中的默认设置。许多简单仿真可以直接使用默认参数。显然，停止时间“Stop time”应该设为至少 5 个时间常数的值。起始时间可以设成合适的值。多数仿真都可以使用默认的仿真算法 ode45（即 Dormand-Prince 公式），这种算法是最佳首选，它在计算下一个状态的输出值时只需用到当前的输出值。

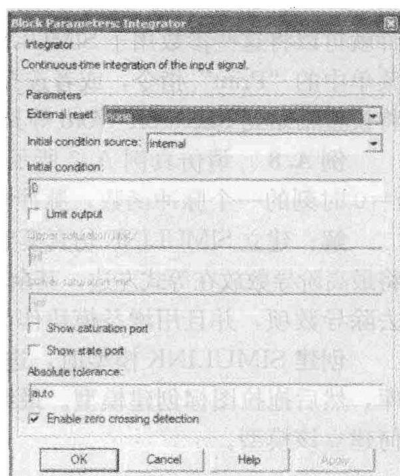
仿真的运行可以用固定步长求解器或者用变步长求解器。固定步长求解器从起始时间开始，采用等间隔时间步长计算。步长要足够小，才能使近似计算有效；但也不能太小，使得仿真运行时间过长。经验规则是：先设定

$$步长=\frac{仿真总时间}{500\sim5000}$$

然后，将步长减小为厘米的 1/10，如果仿真输出没什么变化，就保持该步长；否则再减小为厘米的 1/10，直到输出稳定为止。变步长算法会调整步长，当输出变化较快时，就减小步长；当输出变化较慢时，就增加步长。变步长仿真通常可以节省仿真计算时间，加快仿真速度。选择仿真菜单中的“Start”指令或者单击运行指令的图标，就可以开始运行仿真。

SIMULINK 模型建立之后，选择“File”菜





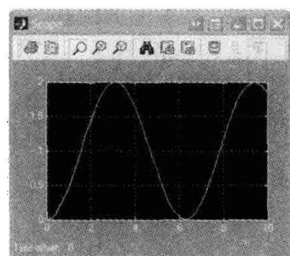


图 A.11　上：SIMULINK 模块图；中：积分器模块的参数；下：示波器模块的输出

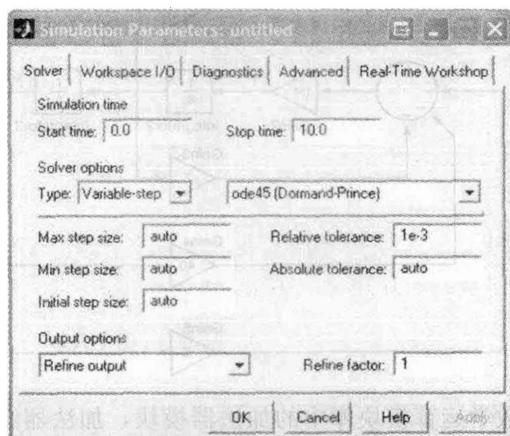


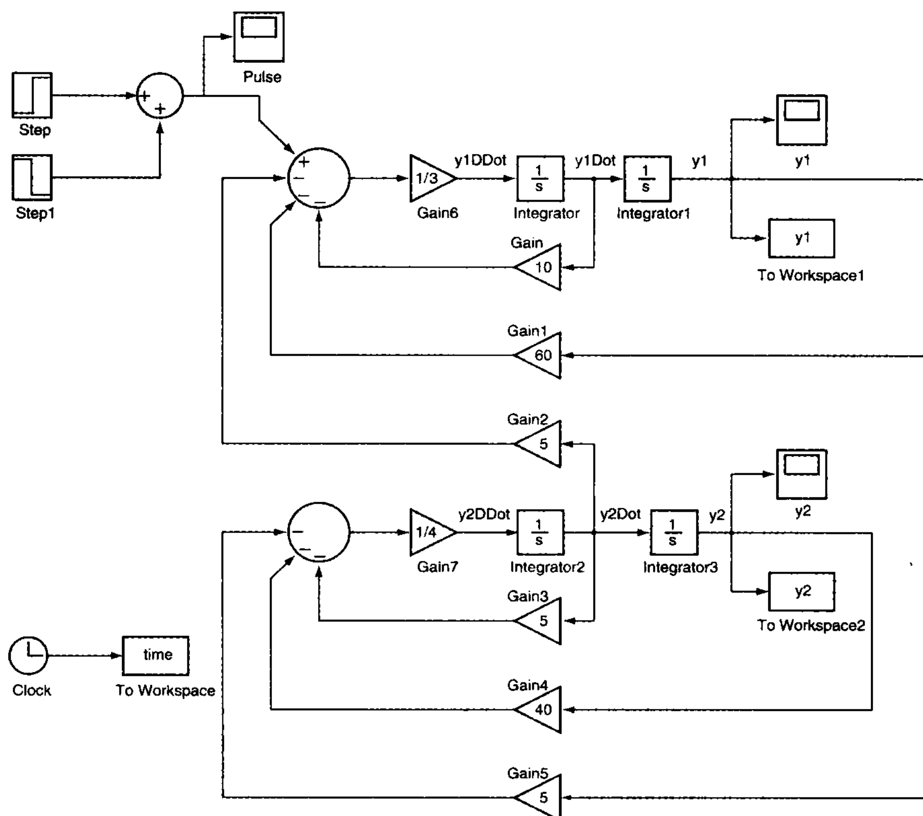
图 A.12　仿真参数窗口

单中的“Save”或者“Save as”选项可以保存模型。一般仿真模型都有一些参数，每次仿真时有些参数的取值会不一样。为了方便，在 MATLAB 中定义参数比较好，例如，$K=5$，$M=3$ 等，这样可以把参数值保存在 m 文件中。在 MATLAB 中打开 m 文件就可以将这些参数用于 SIMULINK 模型。如果要打印 SIMULINK 模型，选择“File”菜单中的“Print”指令，或者选择“Edit”菜单中的“Copy Model to Clipboard”指令并将模型粘贴到 Microsoft Word 等其他应用软件中。

例 A.8 请仿真例 A.7 所示系统的输出响应，假设初始条件为 0，输入 $f(t)$ 为 $t=0$ 时刻的一个脉冲函数，脉冲幅值为 120，宽度为 10。

解： 建立 SIMULINK 仿真模型时最好先画出框图。例 A.7 已经建好了本题的框图：将最高阶导数放在等式左边，其余项放在等式右边；用加法器完成等号功能，用积分器去除导数项，并且用增益模块作为反馈元件。SIMULINK 模型基本上与框图相同。

创建 SIMULINK 模型时，进入 SIMULINK，打开工作空间，并打开相应的模块库，然后拖拉图标创建模型。最终的 SIMULINK 模型如下图所示。下面详细介绍如何建立该模型。



开始时，先添加数学运算模块库中的加法器模块，加法器图标为⊕。双击模块打开模块的参数设置窗口，可以将该模块的默认圆形改变成长方形，也可以通过添

加“+”号或“-”号来增加输入的端口数目。本题的第一个方程有 4 个元件要连接到加法器上，1 个正的输入元件和 3 个负的反馈元件，因此，去掉默认加法器的一个正输入端并添加 3 个负输入端，就变成了。

系统输入是幅值为 120、宽度为 10 的脉冲函数，这里用信号源模块库中的两个阶跃输入模块来创建这个脉冲函数，阶跃输入模块的图标是 Step。双击模块打开模块的参数设置窗口，将第一个模块的阶跃时间“Step Time”设为 0，起始值“Initial value”设为 0，终值“Final value”设为 120。将第二个模块的阶跃时间“Step Time”设为 10，起始值“Initial value”设为 0，终值“Final value”设为-120。再将这两个阶跃输入模块的输出连接到一个新的加法器上，这个加法器的输出连接到第一个加法器的正输入端。

第一个加法器的输出要连接一个 1/3 增益模块。增益模块在数学运算模块库里，其图标是 1 Gain。从模块库中拖出增益模块之后，双击模块打开模块的参数设置窗口，将增益值设为 1/3。如果可以显示，增益模块中就标出增益值，否则就显示“-k-”。

然后，从连续模块库中拖出两个积分器，积分器上标有$\frac{1}{s}$，它表示积分运算的拉普拉斯变换。有时，积分器具有非 0 初始值，本例中并不需要。如果要设置积分器的初始值，就双击模块打开模块的参数设置窗口。积分器的默认初始值为 0。加法器之后的增益模块的输出连接到第一个积分器的输入端，第一个积分器的输出连接到第二个积分器的输入端。最好给每个积分器的输出加上标签。单击工作空间任意空余的地方就可以添加标签。一种较方便的表示方式是将 $\ddot{y}_1$ 标为“y1DDot”，将 $\dot{y}_1$ 标为“y1Dot”，依此类推。

再从数学运算模块库中拖出 3 个增益模块。为了改变增益模块的方向，选中模块，按“Ctrl-R”键两次就可以将其旋转 180°，每按一次旋转 90°。双击模块可以将 3 个增益值分别设定为 10、60 和 5。将积分器的输出连接到增益模块的输入端，按下“Ctrl”键并单击鼠标左键可以将连线分支。各个增益模块的输出连接到加法器上。注意，在第二个方程模型建立之后，$\dot{y}_2$ 项才能连接到增益模块上。

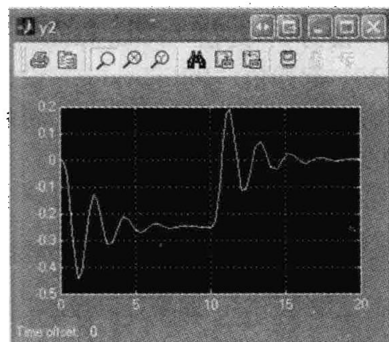
用同样的步骤可以建立第二个方程的模型。注意，Gain2 模块的输入是 $\dot{y}_2$，Gain5 模块的输入是 y_1。

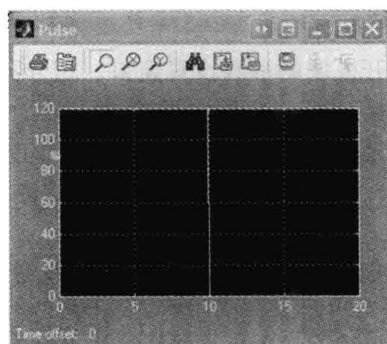
为了显示输出结果，再添加了接收器模块库中的两个示波器模块，并双击名称标签，将它们改为“y1”和“y2”，以便于区分。双击示波器图标可以打开一个新的窗口用于显示输出变量曲线。最好在输入脉冲处也加一个示波器，以便于确认输入信号是否正确。示波器图形窗口菜单上的望远镜图标可用于自动调节图形的大小。

来自信号源模块库的时钟模块用于输出每一步仿真之后的仿真时间。为了将 SIMULINK 的仿真数据传送到 MATLAB 的工作空间作进一步处理和作图显示，还用了接收器模块库的“To Workspace”模块。本例一共用了 3 个“To Workspace”模块，分别传送变量 y_1、y_2 和时间。“To Workspace”模块使用的默认变量名是“si-

mout1”（第一个变量）、“simout2”（第二个变量）等。双击模块可以方便地修改这些变量名，使用 y_1、y_2 和 Time 等具有含义的变量名。另外，保存格式“Save format”应该改为数组“Array”，采样时间“Sample Time”设置成可以提供足够多数据点的值。本例中，仿真时间为 20，因此采样时间设为 0.2，这样，MATLAB 工作空间中的每个变量会有 100 个数据点。数据传送到 MATLAB 工作空间之后，就可以用于处理和显示，例如，用“plot（Time，y1）”指令作图等。







最后一步是打开仿真参数设置窗口，将“Stop time”设为 20，并单击运行按钮，上面这些图显示了 3 个示波器的输出结果。

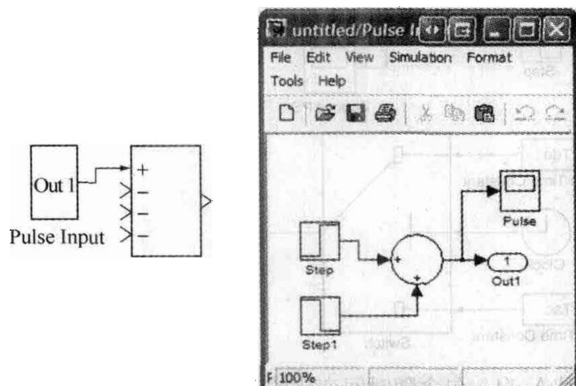
有时，需要采样不同的模型参数进行仿真计算，以便考察模型的性能。本例模型的模块参数要逐个调整。如果模型包含有数百个参数，那么，调整参数的工作量就非常大。此时如果在 MATLAB 中输入参数值，模型仿真只用模块的参数名，而不是参数值，就要方便多了。例如，在 MATLAB 中输入如下信息：

```
≫PulseGain=120；
≫PulseDuration=10；
≫Gain=10；
≫Gain1=60；
≫Gain2=5；
≫Gain3=5；
```

```
≫Gain4=40;
≫Gain5=5;
≫Gain6=1/3;
≫Gain7=1/4;
```

然后，每个增益模块都用参数名，而不是参数值。MATLAB中参数值的任何变动都会自动反映到SIMULINK中。积分器模块初始值的设定也可以在MATLAB中完成。把初始值保存在m文件中，就不需要在每次重新启动MATLAB时输入参数值了。■

建立SIMULINK模型时，如果利用子系统把模块组合起来，可以提高模型图的可读性。例如，本例中可以将脉冲输入画成一个子系统，只要用鼠标拉出一个框，围住相关的几个模块，再单击鼠标右键，选择创建子系统“Create subsystem”即可。子系统的输入和输出就是所选中模块的那些相应信号。图A.13显示了脉冲输入子系统及其与模型其余部分的连接。双击子系统模块可以打开子系统窗口，显示子系统中各个模块，其中子系统的输出用Out1模块表示。



图A.13　左图：脉冲输入子系统；右图：双击子系统打开的新的窗口，显示子系统包含的模块

“Math Function”和“Fcn”这两个模块在仿真中很有用。“Math Function”来自数学运算模块库，用它可以方便地通过下拉菜单输入指数、对数、幂函数等多种数学函数。例如，假设输入变量名为“u”，函数为“exp”，输出为“exp(u)”。模块上的标签就是数学函数。“Fcn”来自用户自定义函数“User-Defined Functions”模块库，用它可以输入各种表达式。假设输入变量名还是“u”，“Fcn”模块可以使用任何加、减、乘、除等代数运算，以及正弦、余弦、指数、幂等数学函数，也可以使用任何MATLAB参数。例如，表达细胞膜钾离子通道的电导要用表达式$\bar{g}_k n^4$，其中$\bar{g}_k$是常数，n是随膜电压和时间变化的速率常数。在SIMULINK中用“Fcn”模块，将其参数设置成“Gkbar * u^4”，其中“Gkbar”的值在MATLAB中定义，“u”是模块的输入，代表n（参见第11章）。

系统仿真时，一些参数的取值可能随着阈值条件而变化。例如，在扫视眼动系统中（参见第12章），主动肌的输入信号F_{ag}是一个经过低通滤波的脉冲一阶跃信

号，可以用如下方程描述：

$$\dot{E}_{ag}=\frac{N_{ag}-F_{ag}}{\tau_{ag}}$$

其中，N_{ag}为神经控制的输入信号（即脉冲一阶跃波），并且

$$\tau_{ag}=\tau_{ac}(u(t)-u(t-t_1))+\tau_{de}u(t-t_1)$$

其中 τ_{ag}是随时间变化的时间常数函数，它由激活时间常数 τ_{ac}和去激活时间常数 τ_{de}构成。如图 A. 14 所示，建立这个子系统时，我们用了信号路由“Signal Routing”模块库中的开关“Switch”模块。并用时钟模块监测仿真时间。仿真开始时，用时间常数 τ_{ac}；当仿真时间大于 $t1$ 时，达到阈值，开关切换到时间常数 τ_{de}。图 A. 14 的其余模块用于求解 F_{ag}。

有关 SIMULINK 的其他功能，请读者参考帮助信息。

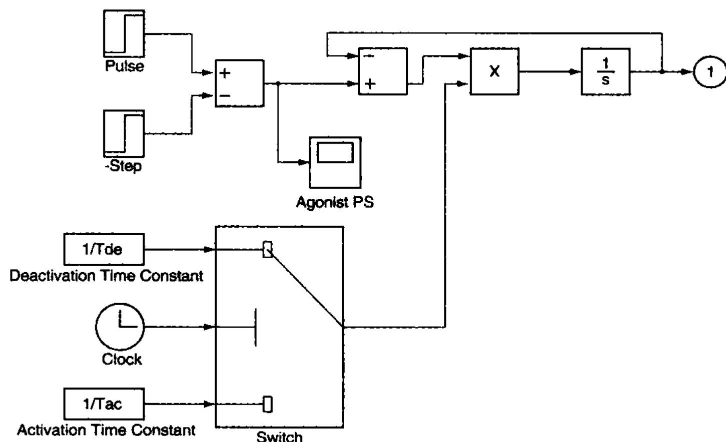


图 A. 14　包含随时间变化的时间常数的低通滤波器

国际视野 科技前沿

国际信息工程先进技术译丛

上架指导：工业技术/生物医学



地址：北京市百万庄大街22号
电话服务
社服务中心：(010)88361066
销售一部：(010)68326294
销售二部：(010)88379649
读者服务部：(010)68993821

邮政编码：100037
网络服务
门户网：http://www.cmpbook.com
教材网：http://www.cmpedu.com
封面无防伪标均为盗版

ISBN 978-7-111-30125-7

封面设计：马精明

定价：98.00元

